Imaging spectrometry-derived estimates of regional ecosystem composition for the Sierra Nevada, California

Article (Accepted Version)

Imaging Spectrometry-derived Estimates of Regional Ecosystem Composition for the Sierra Nevada, California

Highlights

Stacy A. Bogan, Alexander S. Antonarakis, and Paul R. Moorcroft*

- Imaging spectrometry used to estimate fine-scale terrestrial ecosystem composition
- Analysis conducted over an ecologically diverse California landscape
- Sub-pixel PFT abundances accurate to an average RMSE of 11.3% and Bias -2.4 to 3.1%.
- Method provides a framework for improved terrestrial biosphere model simulations

Imaging Spectrometry-derived Estimates of Regional Ecosystem Composition for the Sierra Nevada, California

Stacy A. Bogan[1], Alexander S. Antonarakis[2], and Paul R. Moorcroft[3*]

Affiliations:

[1] Center for Geographic Analysis, Harvard University, 1737 Cambridge Street, Cambridge, Massachusetts 02138 USA.

[2] Department of Geography, Sussex University, Falmer, Brighton, BN1 9QJ, UK

[3] Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts 02138 USA.

*corresponding author: paul_moorcroft@harvard.edu,
ph: +1-617-496-6744

Abstract

The composition of the plant canopy is a key attribute of terrestrial ecosystems, influencing the fluxes of carbon, water, and energy between the land surface and the atmosphere. Terrestrial ecosystem and biosphere models, which are used to predict how ecosystems are expected to respond to changes in climate, atmospheric CO_2, and land-use change, require accurate representations of plant canopy composition at large spatial scales. The ability to accurately specify plant canopy composition is important because it determines the physiological and ecological properties of plants (such as leaf photosynthetic capacity, patterns of plant carbon allocation and tissue turnover, and the resulting dynamics of plant demography) that govern the biophysical and biogeochemical functioning of ecosystems. Traditionally, plant canopy composition has been represented in a coarse-grained manner within terrestrial biosphere models, with ecosystems being comprised of a single plant functional type (PFT). However, models are increasingly seeking to represent fine-scale spatial variation in plant functional diversity. In this study, we show how imaging spectrometry measurements can provide spatially-comprehensive estimates of within-biome heterogeneity in PFT composition across a functionally diverse and topographically heterogeneous ~710 km^2 area in the Southern Sierra Mountains of California. AVIRIS (Airborne Visible Infrared Imaging Spectrometer) data at 18 m resolution from the recent HyspIRI Preparatory Mission (Hyperspectral InfraRed Imager) were used to estimate the sub-pixel fractions of seven PFTs represented in the ED2 terrestrial biosphere model: Shrub, Oak, Western Hardwood, Western Pine, Cedar/Fir, and High-elevation Pine, plus a Grass/NPV (Non-Photosynthetic Vegetation) fraction using Multiple Endmember Spectral Mixture Analysis (MESMA). ED2 is an individual-based terrestrial biosphere model capable of representing fine-scale sub-pixel ecosystem heterogeneity. Our results show that this methodology captures important elevation-related shifts in canopy composition that occur within the study area that are not resolved by existing multi-spectral land-cover products. These estimates modestly improved when the putative PFT endmembers considered in the mixture analysis were constrained using available geospatial data about the presence and absence of the

PFTs in particular areas: the average RMSEs (root-mean-square errors) with the geospatially-constrained versus conventional method were 11.3% and 11.9% respectively, with larger reductions in the bias (i.e. mean error) in the abundances of Oak, Cedar/Fir, and Western Hardwood PFTs (ranging from 2.0% to 7.8%). At the hectare scale around four flux towers in the Southern Sierra Mountains, the overall composition improved from an RMSE of 18.2% (5.0-24.2% for individual PFTs) to RMSE 9.5% (3.3-13.2% for individual PFTs). Downgrading AVIRIS to 30 m resolution resulted in a reduction in accuracy of the constrained method to an RMSE of 12.7% (0-23.7%) with < 1% change in bias for all tree and shrub PFTs. Our results demonstrate that imaging spectrometry measurements from planned satellite missions such as HyspIRI, EnMAP (Environmental Mapping and Analysis Program), and HISUI (Hyper-spectral Imager SUIte) can provide important and much-needed information about fine-scale heterogeneity in the composition of plant canopies for constraining and improving terrestrial ecosystem and biosphere model simulations of regional- and global-scale vegetation dynamics and function.

Keywords: Imaging Spectrometry, Canopy Composition, HyspIRI, Multiple Endmember Spectral Mixture Analysis

1. Introduction

Our understanding of how terrestrial ecosystems are changing, and will continue to change as a result of ongoing changes in climate, increasing atmospheric CO_2 concentrations, and land-use, relies heavily on terrestrial ecosystem and biosphere models. These models incorporate the impact of these different environmental forcings on terrestrial vegetation dynamics, biogeochemistry, and water and energy fluxes (Fisher et al., 2014).

Traditionally, the composition of the plant canopy within terrestrial ecosystem and biosphere models has been represented in a coarse-grained manner, with each of the earth's major biomes being represented as a single, homogeneous Plant Functional Type (PFT) (e.g. C_3 grasses, needle-leaved evergreen trees, or cold- or drought-deciduous trees) (Huntzinger et al., 2012; Fisher et al., 2014). However, results of recent terrestrial biosphere modeling studies (e.g. Levine et al., 2016; Sakschewski et al., 2016) have highlighted the importance of fine-scale diversity and heterogeneity in canopy composition for predictions regarding the responses and resilience of terrestrial ecosystems to climatological perturbations. In these models, the plant canopy within each climatological grid cell is compositionally heterogeneous at the scale of individual plants within the canopy, i.e. at spatial scales of meters (see Fisher et al. 2018 for a recent review).

The incorporation of fine-scale ecosystem heterogeneity in terrestrial biosphere models has, in large part, been motivated by the desire to incorporate a greater diversity of PFTs into model simulations because diversity is an important source of ecosystem resilience (see Moorcroft et al. 2006 and references therein). Building upon the development of regional and

96 global-scale databases of plant attributes, such as TRY (Kattge et al. 2001), GLOPNET (Wright
97 et al. 2004) and the database of Atkin et al. (2015), plant trait databases are increasingly being
98 used to define the characteristics of plants within the terrestrial biosphere models they are using
99 (e.g. Verheijen et al., 2013, Fyllas et al. 2014, Sakschewski et al 2016). Examples of important
100 composition-related parameters within terrestrial biosphere models are given in Table A1: they
101 include leaf photosynthetic properties that govern rates of carbon fixation and transpiration,
102 seasonal patterns of leaf phenology, patterns of plant carbon allocation to leaves, stems and
103 roots, rates of tissue turnover, and resulting rates of plant height and diameter growth and rates of
104 stem mortality.

105 Successful implementation of this new generation of terrestrial biosphere models in
106 regional and global scale modeling studies requires more highly-resolved information about
107 spatial variation in the composition of plant canopies than the information provided by existing
108 global classifications of vegetation types. The ability to more accurately specify plant canopy
109 composition is important because it is used to specify plant physiological and ecological
110 properties that determine the biophysical and biogeochemical functioning of ecosystems being
111 simulated.

112 Remote sensing measurements offer a promising way to provide spatially-resolved,
113 spatially-extensive information about the fine-scale variation in the composition and structure of
114 plant canopies that can be used by the new generation of terrestrial biosphere models to improve
115 their predictions of vegetation dynamics and the associated exchanges of carbon, water and
116 energy between land and the atmosphere. Metrics of fine-scale ecosystem structure, such as
117 forest canopy height (Dubayah & Drake, 2000; Drake et al., 2002; Lefsky et al., 2005) and
118 aboveground biomass (Saatchi et al., 2007; Saatchi et al., 2011; Baccini et al., 2012), have been

119 derived using lidar and radar active remote sensing technologies, and analyses have shown that
120 these measurements can be successfully used to initialize and improve terrestrial ecosystem and
121 biosphere simulations (Hurtt et al., 2004; Hurtt et al., 2010; Antonarakis et al., 2011). Regarding
122 metrics of fine-scale ecosystem composition, Antonarakis et al. (2011) showed that, in addition
123 to measures of ecosystem structure, further improvements to terrestrial ecosystem simulations
124 required spatially-resolved, spatially-extensive information about the fine-scale or sub-pixel
125 variation in the composition of the plant canopy. This is because, as noted earlier, plant
126 composition determines important physiological and ecological properties of the ecosystem,
127 which, in turn, governs its biophysical and biogeochemical functioning.

128 Currently, estimates of regional and global-scale terrestrial canopy composition used in
129 terrestrial ecosystem and biosphere model simulations are derived from multi-spectral remote
130 sensing estimates of broadly-defined land cover categories or global biome types. For example,
131 the MODIS global land cover product (Friedl et al., 2010) distinguishes 8 global biome types, the
132 Global Land Cover Product (GLOBCOVER; Bontemps et al., 2011) distinguishes 18 global
133 vegetation types and the NLCD (National Land Cover Database) Land Cover product
134 distinguishes 9 vegetation types. Figure 1a shows the NLCD (Homer et al., 2015) classification
135 of the region investigated in this study, in which the vegetation in each pixel is assumed to be
136 homogenous and comprised of one of five vegetation types; deciduous, evergreen, mixed forest,
137 shrubs, and grasses. The impact of land cover type uncertainty on terrestrial ecosystem and
138 biosphere model predictions was investigated by Quaife et al. (2008), who showed that incorrect
139 assignment of land cover classes to PFTs, the difficulties in differentiating vegetation types, and
140 information loss at coarse resolutions resulted in significant differences in predictions of
141 terrestrial carbon fluxes.

142 Over the past two decades, studies have shown that the higher spectral resolution of
143 imaging spectrometry sensors can yield more taxonomically- and functionally-resolved estimates
144 of spatial variation in canopy composition (Goodenough et al., 2003; Kokaly et al., 2003; Asner
145 et al., 2008; Féret & Asner 2013; Ferreira et al., 2015; Clark et al., 2018). For example, Martin
146 et al. (1998) classified eleven forest cover types in Harvard Forest, Massachusetts, Kokaly et al.
147 (2003) distinguished 9 forest types within an area of Yellowstone National Park, van Aardt &
148 Wynne (2007) distinguished three pine species in a Virginia forest, and Clark et al. (2018)
149 recently classified twenty-one forest types in the Californian Bay Area.

150 One of the most commonly used methods for estimating vegetation composition from
151 imaging spectrometry measurements is Spectral Mixture Analysis (SMA; Adams et al., 1986),
152 which estimates the fractional contribution of different reference spectral signatures to the
153 reflectance signature of each pixel (Keshava & Mustard, 2002). Compared to other classification
154 methods, such as Random Forests, that assign a single land cover type per pixel, the fractional
155 cover estimates obtained from SMA aligns well with the need to specify fine-scale spatial
156 variation in composition and structure of plant canopies in modern terrestrial ecosystem and
157 biosphere models. See Fassnacht et al. (2016) for a recent review of classification methods
158 applied to forested ecosystems. Ideally, SMA uses the fewest possible endmembers (i.e. 'pure'
159 spectra corresponding to each vegetation or land cover type) to characterize the composition
160 within each pixel (Sabol et al., 1992; Adams and Gillespie, 2006), and requires some prior
161 knowledge of the ecosystem. Multiple Endmember Spectral Mixture Analysis (MESMA;
162 Roberts et al., 1998) builds on traditional SMA by allowing endmembers to vary in type and
163 number on a per pixel basis with combinations of endmembers with the best fit (lowest RMSE)
164 being chosen for each pixel. MESMA also accounts for variation in brightness between

165 endmembers by including a shade endmember, which can be either an image-based endmember
166 or photometric (i.e. spectrally flat, or near-zero reflectance). MESMA has been used for a
167 number of vegetation mapping studies: Roberts et al. (1998) and Hamada et al. (2011) mapped
168 vegetation in arid landscapes of southern California; Li et al. (2005) mapped coastal marsh
169 vegetation in California; and Youngentob et al. (2011) mapped two Eucalyptus sub-genera in
170 Australia.

171 As noted by Clark et al. (2018), however, until recently, the number of studies examining
172 the ability of medium-resolution spaceborne hyperspectral sensors to estimate ecosystem
173 composition has been limited. Three recent studies that have addressed this issue are Roth et al.
174 (2015), who used MESMA to map the dominant plant species in five different ecosystems across
175 North America, Clark et al. (2017) who used MESMA and machine learning to map ten
176 vegetation cover types in the Californian Bay Area, and Clark et al. (2018) who used machine
177 learning to classify twenty-one forest types in the Californian Bay Area. However, in all three of
178 these analyses MESMA was used to identify and validate the dominant cover type within each
179 pixel rather than producing estimates of the fractional composition of vegetation within each
180 pixel.

181 Antonarakis et al. (2014) showed how high spatial resolution (10m) imaging
182 spectrometry-derived estimates of the sub-pixel level fractional PFT composition of an eastern
183 mixed-temperate forest canopy could be combined with lidar-derived estimates of forest canopy
184 structure to provide a spatially-resolved estimate of above-ground plant composition and
185 structure across a ~4 km^2 area of the Harvard Forest Long-Term Ecological Research (LTER)
186 site in the Northeastern United States. Their analysis showed that incorporating this remote
187 sensing-derived estimate of above-ground ecosystem state into an individual-based terrestrial

188 biosphere model simulations yielded marked improvements in the accuracy of the model's
189 predictions of the ecosystem's carbon fluxes compared to those obtained from a conventional
190 "potential vegetation" simulation (i.e. a simulation in which canopy composition is estimated by
191 performing a long-term, multi-century simulation until the vegetation within the model comes
192 into equilibrium with the climatological forcing data). Moreover, the improvements were similar
193 in magnitude to the improvements obtained from simulations initialized with ground-based
194 inventory measurements of canopy composition and structure (Antonarakis et al., 2014).

195 In this analysis, we use imaging spectrometry measurements from the recent National
196 Aeronautics and Space Administration (NASA) Hyperspectral Infrared Imager (HyspIRI)
197 Preparatory Campaign (Hochberg et al., 2015) to estimate sub-pixel level plant functional type
198 composition across a ~710 km^2 area in the Southern Sierra Mountains of California. The purpose
199 of the HyspIRI preparatory campaign is to provide multi-temporal imaging spectrometry data
200 over large areas of California at spatial resolutions of ~18 m and 30 m, i.e. at a spatial resolution
201 approaching the characteristics of anticipated global imaging spectrometry missions, such as
202 HyspIRI (Lee et al., 2015) the Environmental Mapping and Analysis Program (EnMAP, Guanter
203 et al., 2016) and the Hyper-spectral Imager SUIte (HISUI, Matsunaga, et al., 2016) , and
204 NASA's planned Surface Biology and Geology mission (NASEM 2018).

205 The purpose of this study is to evaluate the ability of such instruments to provide reliable,
206 spatially-extensive estimates of fractional PFT composition suitable for constraining individual-
207 based terrestrial ecosystem and biosphere model simulations of large-scale ecosystem dynamics
208 and functioning. As in Antonarakis et al., (2014), key characteristics of the methodology
209 employed here compared to earlier studies are: (i) the direct alignment between definitions of the
210 PFTs whose abundances are being estimated by remote sensing and the PFTs defined within the

terrestrial biosphere model; (ii) use of the sub-pixel PFT fractional abundances to define the fine (meter-scale) sub-grid scale variation in the composition of the plant canopy within the individual-based ecosystem model. The analysis presented here moves beyond that of Antonarakis et al. (2014) in three important respects with regard to assessing the relevance of a global imaging spectrometry mission for constraining regional-to-global scale terrestrial biosphere model simulations. First, we assess the ability of this methodology to work with imaging spectrometry measurements collected at spatial resolutions of planned global missions and to be successfully applied at large spatial scales and over heterogeneous landscapes. Second, we investigate the ability to successfully estimate ecosystem composition in regions of complex terrain that have significant topographic heterogeneity. Third, we evaluate the ability to successfully estimate plant canopy composition across a variety of ecosystem types, including: mixed tree-grass savannahs, deciduous woodlands, and coniferous forests.

2. Materials & Methods

2.1 Study area and Data

The study area is a 710 km^2 area (~67.5 km x 10.5 km) in the Sierra Nevada Mountains northwest of Fresno, California (Figure 1 a-c) (National Elevation Dataset, 2009). It spans an elevational gradient ranging from 125 – 3175 meters above sea level: the western third lies in the foothills of the Sierra Nevada and is dominated by oak savanna ecosystems; the central and eastern portions are higher elevation montane to subalpine ecosystems, moving eastwards from oak-savannas to pine/oak forest, through mixed conifer forests, and then to sub-alpine forest.

The study area encompasses the National Science Foundation's Southern Sierra Critical Zone Observatory (CZO). The CZO includes four eddy-flux tower sites where carbon, water and energy fluxes are being measured and ground-based inventory measurements of canopy

composition and structure are available within 200m x 50m (1 ha) plots located in the footprint of the flux-towers (Kelly, 2014). The four flux tower sites: San Joaquin Experimental Range (SJER), Soaproot Saddle, P301 and Shorthair, span an elevation-related and east-to-west physiographic gradient and accompanying gradient in ecosystem composition (Figure 1b). Further details on the physical and vegetation attributes of the four sites can be found in Table 1.

Table 1. Descriptions of the Four Southern Sierra Critical Zone Observatory Flux Towers.

Flux Tower	Tower Position (Latitude Longitude)	Elevation (m)	Mean Min Temp (°C)	Mean Max Temp (°C)	Precipitation (mm/yr)	Dominant Species
San Joaquin Experimental Range	37.1087 -119.7313	405	9.3	23.5	510	*Quercus douglasii*, *Quercus wislenzni*, *Pinus sabiniana*, *Ceanothus spp., Arctostaphylos spp.*
Soaproot Saddle	37.0311 -119.2563	1160	5.5	18	805	*Pinus ponderosa, Calocedrus decurrens, Quercus chrysolepis, Quercus kelloggii*
P301	37.0673 -119.1948	2015	2.7	14.8	1015	*Abies concolor, Calocedrus decurrens, Pinus lambertiana, Quercus kelloggii, Pinus jeffreyi*
Shorthair	37.0671 -118.9871	2700	-1.9	10.2	1078	*Pinus contorta, Pinus monticola, Abies magnifica*

Geospatial data on the plant species found within the study region were accessed in May 2014 from the United States Department of Agriculture Forest Service CALVEG (Classification and Assessment with Landsat of Visible Ecological Groupings) GIS database (CALVEG: Existing Vegetation, 2009a,b,c; 2010) (Figure 2b). The CALVEG dataset is comprised of polygons (average polygon size 0.026 km^2) categorized into different kinds of vegetation types

245 (referred to as plant associations), produced by the US Forest Service using a combination of
246 Landsat, field verification and expert guidance. Milliken et al. (1998) assessed the accuracy of
247 CALVEG and found that its accuracy of its conifer classes in three areas of Northern California
248 were 73%, 77% and 82% respectively, and the accuracy of its shrub classes was 91% (Milliken
249 et al., 1998). Each plant association contains a list of plant species, but no information is
250 provided about their relative abundances. The polygons spanning our study area were last
251 updated in either 2001 or 2007.

252 Eleven AVIRIS flight-lines were acquired on June 12, 2013 between the times of 17:46
253 UTC and 19:29 UTC, and eleven AVIRIS flight-lines were acquired over the same geographic
254 area on November 05, 2013 between the times of 18:38 UTC and 21:60 UTC as a part of the
255 HyspIRI Preparatory Campaign. The HyspIRI preparatory campaign aims to provide multi-
256 temporal data over large areas of California, to prototype the science that would be enabled by a
257 corresponding global, satellite-based mission (Thompson et al., 2015). The AVIRIS data has
258 224 bands ranging from 400 – 2500 nm (10 nm spectral resolution) with 18 m spatial resolution
259 (Green et al., 1998). Calibrated AVIRIS radiance data was converted to surface reflectance at
260 the Jet Propulsion Laboratory (JPL, Pasadena, California, USA) using a three-step method. First,
261 radiometric corrections were used to convert raw digital numbers into units of radiance, then
262 ortho-rectification was used to produce geo-referenced images (available at multiple spatial
263 resolutions) and finally, atmospheric correction was performed using the Atmospheric REMoval
264 (ATREM) algorithm (Thompson et al., 2015). The ATREM algorithm incorporates the effects of
265 both water vapor content and topographic elevation on optical depth in its calculation of surface
266 reflectance values.
267

268 *2.2 Definition of Plant Functional Types*

269 As noted in the Introduction, terrestrial ecosystem and biosphere models do not typically
270 represent individual species, but instead group species into a series of Plant Functional Types
271 (PFTs) that share similar biophysical, physiological, and ecological properties. In this analysis,
272 we focused on estimating the relative abundance of seven PFTs represented in the Ecosystem
273 Demography model version 2 (ED2) terrestrial biosphere model that are found within the study
274 region: Western Pine, Cedar/Fir, High-elevation Pine, Western Hardwood, Oak, Shrub and
275 Grass. Within the study area, low-elevation grasses are not actively growing in early November
276 and June (AVIRIS images) and for this reason this cover type was labelled Grass/NPV (Non-
277 Photosynthetic Vegetation). Note however, that high-elevation grasses may be active in the
278 summer months (Goulden et al., 2012). The species associated with each of the PFTs is given in
279 Table 2, and the ecological attributes (leaf, root, allocation and allometric traits) of each of the
280 PFTs within the ED2 model are given in Table A1. Each CALVEG Association was classified
281 according to its dominant species using the vegetation description for each Association and from
282 this and the species—PFT correspondences given in Table 2, PFTs were assigned to each
283 CALVEG Association using Table A1.

Table *2*. Species comprising the seven plant functional types (PFTs) distinguished in this analysis.

WESTERN PINE	CEDAR/FIR	HIGH-ELEVATION PINE	WESTERN HARDWOOD	OAK	SHRUB	GRASS/ NPV
Ponderosa Pine (*Pinus ponderosa*	Incense-Cedar (*Calocedrus decurrens)*	Lodgepole Pine (*Pinus contorta*)	Bigleaf Maple (*Acer macrophyllum*)	Blue Oak (*Quercus douglasii)*	Manzanita (*Arctostaphylos spp.*)	Meadow Grass
Western Pine (*Pinus lambertiana*)	White Fir (*Abies concolor*)	Whitebark Pine (*Pinus albicaulis*)	California Sycamore (*Platanus racemosa*)	Black Oak (*Quercus velutina & Quercus kelloggii)*	Ceanothus (*Ceanothus spp.*)	Valley Grass
Knobcone Pine (*Pinus attenuata)*						
Foothill Pine (*Pinus sabiniana)*	Red Fir (*Abies magnifica*)	Western Juniper (*Juniperus occidentalis*)	California Buckeye (*Aesculus californica*)	Interior Live Oak (*Quercus wislizeni)*	Shrub Willow (*Salix spp.*)	
	Douglas-Fir (*Pseudo-tsuga menziesii*)	Single-leaf Pinyon Pine (*Pinus monophylla*)	Mountain Mahagony (*Cercocarpus*)	Canyon Live Oak (*Quercus chrysolepis)*	Mountain Sagebrush (*Artemisia spp.*)	
		Western White Pine (*Pinus monticola*)	Fremont Cottonwood (*Populus fremontii*)		Chamise (*Adenostoma fasciculatum*)	
		Jeffrey Pine (*Pinus jeffreyi*)	White Alder (*Alnus rhombifolia*)			

284 *2.3 The MESMA Algorithm*

285 The relative abundance of the different PFTs was determined from AVIRIS imagery
286 using Roberts et al.'s (1998) MESMA algorithm. MESMA allows endmembers of individual
287 land covers or classification layers to vary on a per pixel basis, resulting in a selection of a
288 weighted combination of endmembers that best describe the reflectance of each pixel (Roberts et
289 al., 1998). The MESMA equation is calculated as:

$$R(\lambda)_i = \sum_{j=1}^{m} f_i^{(j)} \times R(\lambda)_i^{(j)} + \varepsilon(\lambda)_i, \quad \text{and} \quad \sum_{j=1}^{m} f_i^{(j)} = 1, \qquad \text{Eq. (1)}$$

where $R(\lambda)_i$ is the spectral mixture of pixel i in wavelength λ, modeled as the sum of a series of reflectance endmembers, $R(\lambda)_i^{(j)}$ $(j=1...m)$ for pixel i, where m is the number of endmember classes, $f_i^{(j)}$ is the fractional abundance of endmember class j in pixel i, and $\varepsilon(\lambda)_i$ is a residual term describing the unexplained variation in the pixel's reflectance in wavelength λ.

The reflectance endmembers for the m endmember classes in pixel i ($R(\lambda)^{(j)}$, $j=1...m$) are drawn from a spectral library $L(\lambda)^{(j,k)}$ where $k = 1...n_j$ is the number of spectra in endmember class j. MESMA iteratively computes a series of linear mixture models using the n_j different endmembers for each endmember class j, including a shade endmember to account for variation in illumination. In doing so, MESMA accounts for the observed variability in the endmembers of each endmember class. The preferred model for a given pixel i is the one that minimizes the RMSE over all spectral bands:

$$RMSE = \left[\sum_{\lambda=1}^{B} (\varepsilon(\lambda)_i)^2 / B \right]^{1/2}, \qquad \text{Eq. (2)}$$

where B is the number of spectral bands. Following Roberts et al. (1998), the endmember combination with minimum RMSE was selected, there were no constraints on the maximum RMSE, and the shade fraction ranges from 0-1.

The analysis was conducted using VIPER Tools version 2.0. VIPER Tools 2.0 is an IDL-based ENVI (Harris Geospatial Solutions) extension that can be used to create, organize and evaluate spectral libraries, execute MESMA, and shade-normalize MESMA results. It builds upon VIPER Tools 1.5 (Roberts et al., 2007), but has several improvements to the MESMA algorithm. In particular, with regard to the analyses conducted here, VIPER Tools 2.0

311 (https://sites.google.com/site/ucsbviperlab/viper-tools) does not allow negative shade fractions,
312 thereby ensuring that shade fraction estimates are physically reasonable.
313
314 *2.4 Image Processing*
315 Individual AVIRIS flight-lines from the spring and fall flights were mosaicked into a
316 single spring image and a single fall image in ArcGIS. In areas of overlap between flight lines,
317 the pixel with the highest reflectance value for each band was used. The spring and fall images
318 were then subset to the study area (the red region in Figure 1d). Spectral bands that had high
319 levels of atmospheric contamination and bands with low signal-to-noise ratio were removed,
320 reducing the number of bands in each image from 224 to 167 bands covering three areas of the
321 spectrum: 443 – 1342 nm, 1422 – 1790 nm, and 2106 – 2405 nm. A small number of pixels (less
322 than 0.005%) that had anomalously high (above 100%) or extremely low (below 0%) reflectance
323 values were replaced by values linearly interpolated from the reflectance values in the adjacent
324 wavelengths. The two images were then stacked into a single composite image with 167 x 2 =
325 334 spectral bands. Finally, areas of the image that were urban, agriculture, barren, or water
326 were masked out using the CALVEG data layer. A 30 m resolution version of the composite
327 AVIRIS image was then produced by resampling the native 18 m to 30 m resolution using
328 bilinear spatial resampling to evaluate the accuracy of the PFT classification methodology at the
329 resolution of the expected future HyspIRI mission.
330
331 *2.5 Development and Analysis of the Spectral Library*
332 An endmember spectral library was then created by taking targeted samples from the
333 image for the dominant species associated with each of the PFTs (See Table 3). Samples were

first chosen using ground knowledge of where dominant species associated with each PFT in Table 3 are located, or by identifying CALVEG polygons whose associations were dominated by species associated with, or where the target species existed with other easily distinguishable species (e.g. Oak and Shrubs). Areas that, according to CALVEG, were highly mixed, such as the Mixed Conifer Pine were avoided (see Figure 2b). High-resolution (30–60 cm) imagery in ArcGIS (sources: ESRI, DigitalGlobe, GeoEye, i-cubed, USDA FSA, USGS, AEX, Getmappint, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community) was then used to identify areas that were densely vegetated and contained high canopy cover of the target species. In addition to the vegetation signatures, a Bare-ground class, consisting of rock and soil spectral signatures, was also defined. Signatures for this class were sampled either in mountainous areas (for exposed rock) or in areas that appeared in the high-resolution imagery to be bare soil. False color composites of both the June and November images were used to ensure sampling vegetated or non-vegetated areas, and to avoid sampling overly-shaded areas, areas impacted by BDRF affects, and areas with snow cover. Samples consisted of two to several pixels, and endmembers were created by averaging the signatures of all pixels in a sample. Further details on the endmember samples can be found in Table 3.

Each pixel in the full transect study area was assumed to be comprised of up to three of the eight endmember classes (7 PFTs plus Bare-ground) plus photometric shade in the MESMA algorithm in VIPER tools. Accordingly, each pixel was tested for: each signature in one class + shade (i.e. a 2 endmember model); each of the signature combinations between two classes + shade (i.e. a 3 endmember model); and each of the signature combinations between three classes + shade (i.e. 4 a endmember model). The selection of the best fit model for each pixel was based

356 on the smallest RMSE (see Equation 2); models with more signatures were not weighted

Table 3. Details of the image areas sampled for each Plant Functional Type (PFT) endmember class and the Bare-ground endmember class: columns show the area sampled per PFT, the number of samples per PFT, the minimum and maximum latitude and longitude of the areas sampled, and the elevation range of the samples.

PFT plus Background)	Area Sampled (m^2)	Number of Samples	Sample Max. Lat. (°N)	Sample Min. Lat. (°N)	Sample Max. Long. (°E)	Sample Min. Long. (°E)	Sample Elevation range (m)
Western Pine	48,600	11	37.899	36.542	118.955	120.097	217 – 1462
Cedar/Fir	42,444	11	37.763	36.91	119.030	119.81	1327 – 2517
High Pine	36,288	18	38.221	36.993	118.697	119.744	2059 – 3237
Western Hdwd	85,536	23	38.106	36.657	118.629	120.07	109 – 2722
Oak	24,624	10	37.116	37.036	119.324	119.688	232 – 1679
Shrub	57,996	26	38.138	36.957	118.803	119.92	285 – 2921
Grass/NPV	49,896	9	37.121	37.048	119.053	119.738	204 – 2615
Bare-ground	43,094	11	37.996	37.029	118.982	120.021	101 – 3321

357 differently compared to models with less.

358

359 We evaluated two approaches in choosing which PFTs to include for each pixel. In the
360 first approach, hereafter referred to as *conventional*-MESMA, all signature combinations of all
361 eight endmember classes listed in Table 3 (the seven PFTs plus Bare-ground) were evaluated on
362 every pixel within the study area. In the second approach, hereafter referred to as *constrained*-
363 MESMA, an endmember reduction/pruning method was applied as defined in previous studies
364 (e.g. Garcia-Haro et al. 2005; Liu and Yang 2013 Roth et al. 2015). In constrained-MESMA,
365 only signatures from PFTs in Table 3 that the CALVEG habitat map indicated were likely to be
366 present, plus Bare-ground, were tested in each pixel. For example, if a pixel belonged to the
367 CALVEG Mixed Conifer-Fir association (described as primarily containing White Fir, Red Fir,
368 Western White Pine, Lodgepole Pine, Ponderosa Pine, Jeffrey Pine, Manzanita and Ceanothus),

it was evaluated only with spectra for the PFTs containing the species listed in that association (i.e. Cedar/Fir, High-elevation Pine, Western Pine, and Shrub, see Table 2 and Table A2) plus the Bare-ground endmember class. Following the MESMA classification, the PFT abundances within each pixel were re-normalized to account for the shade fraction, which was assumed to be distributed evenly among the endmember classes (the PFTs considered plus Bare-ground) identified within each pixel.

The spectral separability between all pairs of PFTs was calculated using the Jeffries-Matusita Distance (JM Distance), a widely used separability criterion for the evaluation of classification results. This metric is included in this study to understand separability and similarity of endmember spectra used in determining PFT abundances derived from MESMA. Signatures from a pair of PFTs that are spectrally separate will most likely result in more distinct classifications and higher classification accuracy, while spectrally similar signatures may result in pixels classified with incorrect overlapping PFTs. The JM distance equation is calculated as:

$$J_{xy} = 2(1 - e^{-Q}), \quad \text{Eq. (3)}$$

where J_{xy} is the spectral separability of PFTs x and y, and the exponent Q is given by:

$$Q = \frac{1}{8}(\overline{\boldsymbol{x}} - \overline{\boldsymbol{y}})^t \left(\frac{\boldsymbol{\Sigma_x} + \boldsymbol{\Sigma_y}}{2}\right)^{-1} (\overline{\boldsymbol{x}} - \overline{\boldsymbol{y}}) + \frac{1}{2}\ln\left(\frac{\left|\frac{\boldsymbol{\Sigma_x}+\boldsymbol{\Sigma_y}}{2}\right|}{|\boldsymbol{\Sigma_x}|^{\frac{1}{2}}\,|\boldsymbol{\Sigma_y}|^{\frac{1}{2}}}\right), \quad \text{Eq. (4)}$$

where $\overline{\boldsymbol{x}}$ and $\overline{\boldsymbol{y}}$ are, respectively, the mean spectral signature vector for PFTs x and y, t is the transpose function, and $\boldsymbol{\Sigma}_x$, $\boldsymbol{\Sigma}_y$ are the covariance matrices of PFT x's and PFT y's spectral signature vectors (Richards and Jia, 2006).

2.6 Evaluation of the MESMA Classification

MESMA classifications were first evaluated at the four flux towers shown in Figure 1. Ground-based estimates of the tree PFT composition at the four flux tower sites were calculated

392 from forest inventory measurements collected in a 1-hectare (200 m x 50 m) plot at each of the
393 four sites. The plots were oriented along the dominant daytime wind direction at each site, i.e.
394 aligning with the general orientation of the flux-tower footprints. The diameter (units, cm) and
395 species identity of each stem within each plot was recorded (Kelly, 2014). The fractional
396 contribution of each tree PFT to the tree canopy composition was then estimated by aggregating
397 the cross-sectional basal area of the individual trees to obtain the total basal area (units, cm^2 m^{-2})
398 for each of the five tree PFTs and then dividing by the total tree basal area with the plot, under
399 the commonly-used and well-supported assumption that a tree's canopy area is strongly
400 correlated to its basal area (e.g. Cade, 1997 for Western US forests).

401 Ground-based inventory measurements from ninety-five plots collected at low- to mid-
402 elevations were then used for a subsequent independent validation of the sub-pixel MESMA-
403 derived canopy composition estimates. Between 2014-2015, 76 plots were collected by Dr.
404 Huesca-Martinez and Professor Susan Ustin (University of California, Davis), measuring
405 individual tree positions, tree height, crown diameter, species, and trunk diameters of trees > 2
406 cm within 14 - 100 m^2 plots and 62 - 400 m^2 plots. In 2013, 19 plots were collected by the
407 CZO/NEON project, measuring again individual tree positions, tree height, crown diameter,
408 species, and trunk diameters of trees > 0.2 cm within variable size plots of 200-1500 m^2. The
409 plots covered all principal species over the 710 km^2 study area, except high elevation pines, and
410 only 56 of the total 95 plots measured shrub species.

411 Each MESMA-classified pixel is a horizontal representation of PFT abundance. The
412 validation dataset was comprised of the above plot measurements of crown diameters, heights
413 and tree locations, and was converted to estimate of canopy cover for each PFT in each plot.
414 Each tree in a plot, starting from the shortest to tallest, was assigned a circular crown according

to its crown diameter, with taller trees overlapping shorter crowns. The area of crowns belonging to each PFT in a plot was then calculated and divided by the plot area to obtain fractional abundance, with crown segments crossing the boundary of a plot excluded. When evaluating a MESMA-derived canopy composition estimate against the plot measurements a multidirectional buffer of 10 m was considered from the center of each field plot, to account for ~8-10 m GPS accuracy in positioning of individual trees, and the 10% (i.e. 1.8m) AVIRIS spatial calibration accuracy. Since the GPS errors were random with respect to direction, composition estimates were calculated for the pixel containing the reported GPS location of the plot and the surrounding eight pixels, and the estimate with the lowest RMSE used. Plot-level validation using the pixel containing the reported location are given in Table A3.

3. Results

3.1 Conventional MESMA Classification

The PFT composition of the CZO Study Area estimated by conventional MESMA is shown in Figure 2c. The analysis indicates that the western, lower elevation (<875 m) portion of the study area is comprised of Mediterranean savannah ecosystems dominated by grasses interspersed with oaks, shrubs and western hardwoods. At mid-elevation sites (875-1700 m), located in the center portion of the study area, canopy composition changes to a mixture of Shrub-dominated, Oak-dominated, and Cedar/Fir dominated ecosystems (indicated by the orange, red and blue colors respectively in Figure 2c). At the higher elevation (>2000 m) sites, on the eastern portion of the study area, the canopy composition changes to primarily Cedar/Fir or High-elevation Pine dominated ecosystems (indicated by the blue and light green colors respectively in Figure 2c).

437 The corresponding fractional abundance estimates of the seven PFTs obtained from the
438 conventional MESMA approach are shown in Figure 3. Comparing these spatial abundance
439 patterns to the corresponding spatial pattern of the PFTs indicated by their presence and absence
440 in CALVEG habitat classification (Figure A1) indicates three principal differences. First,
441 conventional MESMA estimates that the Grass/NPV, Shrub and Oak PFTs are present at low
442 abundances at high elevation areas on the eastern side of the study domain, while CALVEG
443 indicates that they are absent in these areas (Figure A1, Grass/NPV: compare panels (a) and (c)
444 respectively; Shrub: compare panels (d) and (f) respectively; Oak: compare panels (g) and (i)
445 respectively). Second, conventional MESMA estimates that the Western Hardwood PFT is
446 broadly distributed across the study area while CALVEG indicates that it is restricted to the
447 western portion of the study area (Figure A1, compare panels (j) and (l) respectively). Third,
448 conventional MESMA estimates that the Cedar/Fir and High-elevation Pine PFTs occur at low
449 abundances in low-elevation areas in the western side of the study area while CALVEG indicates
450 that they are absent from these areas (Figure A1, Cedar/Fir: compare panels (p) and (g)
451 respectively; High-elevation Pine: compare panels (s) and (u) respectively).

452 The fractional abundance estimates of the seven PFTs also reveals noticeable banding in
453 the fractional abundances of the Oak and Western Hardwood PFTs at the intersection of the two
454 eastern-most flight lines in the study area (Figure 3, panels c and d). As shown in Figure A2, this
455 reflects differences in the Bidirectional Reflectance Distribution Functions (BRDFs) of pixels in
456 different flight lines affecting the MESMA composition estimates.

457 Figure 4 shows conventional MESMA abundance estimates and corresponding ground-
458 based 1-ha inventory estimates of composition available in the footprint of the four CZO flux
459 tower sites in the study area. At the low-elevation SJER flux tower site, conventional MESMA

460 estimates the abundances of the Oak and Western Pine PFTs at 44% and 25% compared to
461 ground-based inventory estimates of 84% and 16%, and estimates a significant fraction (31%) of
462 the Cedar/Fir PFT that was not present in the ground-based inventory. At the Soaproot flux tower
463 site, conventional MESMA captures reasonably well the observed fractions of the Western Pine,
464 Oak, and Cedar/Fir PFTs (estimates of 54%, 9%, and 34% respectively, compared to the
465 observed estimates of 57%, 11%, and 32% respectively). However, it also incorrectly estimates a
466 small fraction (1%) the High-elevation Pine PFT that was not present in the ground-based
467 inventory. At the P301 flux tower site, conventional MESMA substantially overestimates the
468 fraction of the High-elevation Pine PFT (estimated 39% compared to 3% observed), and
469 underestimates the fraction of the Cedar/Fir PFT (45% compared to the ground-based inventory
470 measurement of 83%). At the high-elevation Shorthair flux-tower, conventional MESMA
471 estimates match the observed fractions of both the Cedar/Fir PFT (2% compared to 1%
472 observed) and the High-elevation Pine PFT (97% compared to 99% observed) well, but the Oak
473 PFT was also estimated to be present, albeit at very low abundance (1%). The conventional
474 MESMA resulted in an overall accuracy at the flux towers of RMSE 18.2%, with RMSEs of
475 24.2%, 17.7%, 20.0%, 5.0% for Cedar/Fir, High Pine, Oak, and Western Pine respectively.
476 The evaluation of the conventional MESMA product against the ninety-five ground-
477 based inventory plots is shown in Figure 5 and Table 4. Overall RMSEs and biases are shown for
478 woody PFT species with RMSEs of 13.7-19.4% for Western Pine, Western Hardwood, Oak and
479 Cedar/Fir, and RMSEs < 5% for Shrubs and High Pines. Note that none of the plots contained
480 High Pine, so non-zero values for this PFT indicate commission errors. Small biases were found
481 for Western Pine (0.6%), Shrubs (0.9%), and High Pine (0.4%), with larger biases for Cedar/ Fir
482 (-4.4%), Oak (-7.1%), and Western Hardwood (8.3%). An assessment of the plot-level accuracy

is shown in Figure A3, which compares the MESMA derived estimates of canopy composition against the ground-based plot measurements of canopy composition. Grass/NPV and Bare-ground fractional abundances were not explicitly evaluated from the 95 plots, as their fractional cover were unknown. Validation using endmember background and grass pixels produced RMSEs of 34.1% and 55.4% (14.7 and -34.5% bias), with a combined Grass/NPV+Bare-ground RMSE of 34.5% and bias of -19.9%.

Table 4. Plot-level Validation of the Conventional 18 m and Constrained 18 m and downscaled 30 m MESMA methods using ninety-five ground-based inventory plots. MESMA classifications are compared to the horizontal crown area fraction occupied by each PFT in a field plot. * None of the 95 plots contained any High Pine, so non-zero values resulting from Conventional MESMA are commission errors.

Plant Functional Type	RMSE – Conventional 18m	BIAS – Conventional 18m	RMSE – Constrained 18m	BIAS – Constrained 18m	RMSE – Constrained 30m	BIAS – Constrained 30m
Western Pine	16.1%	0.6%	18.2%	3.1%	19.5%	2.4%
Cedar/Fir	19.4%	-4.4%	22.2%	-2.4%	23.7%	-1.4%
Western Hdwd	13.7%	8.3%	2.6%	0.5%	4.6%	1.0%
Oak	16.0%	-7.1%	16.2%	-1.2%	15.2%	-1.9%
Shrub	4.3%	0.9%	8.9%	1.9%	13.1%	2.5%
High Pine*	(1.8%)	(0.4%)	(0%)	(0%)	(0%)	(0%)

3.2 Constrained-MESMA

The dominant patterns of ecosystem spatial distribution and transition across the CZO Study Area estimated by the constrained MESMA approach are similar to the conventional MESMA (compare Figure 2c and 2d respectively). However, the relative abundances of the dominant PFTs varies between the two methods. In particular, constrained MESMA estimates a higher occurrence of Oak PFT-dominated and lower frequency of Shrub PFT-dominated areas (red versus orange colors respectively in the center-left portion of the domain Figures 2c and 2d), and a higher frequency of Cedar/Fir PFT compared to Western Pine PFT-dominated areas (blue versus dark green colors respectively in the center-right portion of the domain in Figures 2c and 2d). The constrained MESMA approach also results in the Grass/NPV and Oak PFTs being

absent from the high-elevation sites (Figure 6 panels a and c), the Western Hardwood PFT restricted to the western portion of the domain (Figure 6 panel d), and Cedar/Fir and High-elevation Pine PFTs absent from the western low-elevation portion of the study (Figure 6 panels f and g). It also results in higher abundances of the Shrub and Oak PFTs at the mid-elevation sites (see panels b and c of Figures 3 and 5), and increased Cedar/Fir and High-elevation Pine PFT abundances at high elevations (see panels f and g of Figures 3 and 6).

Constrained MESMA abundance estimates compared to ground-based inventory estimates of composition in the footprint of the four flux tower sites are also shown in Figure 4. At the low-elevation SJER flux tower site, constrained MESMA gave a more accurate estimate for the abundance of the Western Pine PFT (20% compared to the ground-based inventory 16%), and the Oak PFT (80% compared to the ground-based inventory 84%). At the Soaproot flux tower site, constrained MESMA captured the observed fractions of the Western Pine, Oak, and Cedar/Fir PFTs (estimates of 51%, 9%, and 40% respectively) compared to observed estimates (57%, 11%, and 32% respectively). At the P301 flux tower site, constrained MESMA overestimates the fraction of the High-elevation Pine PFT (estimated 22% compared to 3% observed), and underestimates the fraction of the Cedar/Fir PFT (65% compared to the ground-based inventory measurement of 83%). At the high-elevation Shorthair flux tower site, constrained MESMA overestimates the Cedar/Fir PFT (18% compared to 1% observed) and underestimates the High-elevation Pine PFT (82% compared to 99% observed). The constrained MESMA resulted in an overall accuracy at the flux towers of RMSE 9.5%, with RMSEs of 13.2%, 12.6%, 3.3%, 4.1% for Cedar/Fir, High Pine, Oak, and Western Pine respectively. This is a modest improvement over the conventional MESMA, which has an overall RMSE of 18.2%

522 (RMSEs of 24.2%, 17.7%, 20.0%, 5.0% for Cedar/Fir, High Pine, Oak, and Western Pine
523 respectively).

524 The differences in the spatial distributions of the PFTs estimated by the conventional and
525 constrained MESMA approaches are also evident in the elevational distribution of the PFT
526 (Figure 7). Using the constrained MESMA method eliminates High-elevation Pine and Cedar/Fir
527 PFTs along the lower (<875 m) elevation areas, with a corresponding increase in the abundance
528 of the Grass/NPV and Oak PFTs in these areas (Figure 7, yellow and red bars respectively).
529 Furthermore, constrained MESMA resulted in the Grass/NPV, Western Hardwood, and Oak
530 PFTs no longer being present at mid- to high- elevations (upper range limits of 1625 m, 1625 m
531 and 2125 m respectively). At mid-elevations (875-2125 m), the application of the geospatial
532 distribution constraint increases the abundance of the Oak PFT. At high elevations (>2125 m) the
533 constrained MESMA method increases the prevalence of the High-elevation Pine and Cedar/Fir
534 PFTs, decreases the abundance of the Shrub PFT, and increases the Bare-ground fraction
535 (indicated by the height of the grey bars in Figure 7).

536 Figure 5 and Table 4 also show plot level validation of the constrained MESMA method
537 against the ninety-five ground-based inventory plots using both 18 m AVIRIS and AVIRIS
538 downscaled 30 m. The constrained MESMA method resulted in Western Hardwood and Shrub
539 RMSEs of < 10%, Western Pine and Oak RMSEs between 10-20%, and Cedar/-Fir RMSE of
540 22.2%. Biases for these PFTs are between -2.4 and 3.1%. The overall improvement in RMSE
541 with the constrained vs. conventional method was modest declining from 11.9% to 11.3%. This
542 reflects reductions in the RMSEs of the Western Hardwood and High Pine PFTs of 11.1% and
543 1.8% respectively combined with increases in the RMSEs of the other PFTs by 0.2-4.6% (see
544 Figure 5 & Table 4). The Constrained 18 m product substantially reduced the biases in the

545 estimates of Western Hardwoods by 7.8% and Oaks by 5.9%. It also slightly reduced the biases
546 of Cedar Fir and High Pine by 2.0% and 0.4% respectively, and slightly increased the biases of
547 Western Pine and Shrub by 2.5% and 1% respectively. The accuracies of the Constrained and
548 Conventional MESMA methods compared to plot-based measurements of composition (Figure
549 A3), show modest improvements in predicted vs observed slopes towards the 1:1 line for Oak
550 and Western Pine PFTs, and eliminates the erroneous occurrence of Western Hardwoods;
551 however, it also decreases in accuracy of the predicted versus observed slope and the
552 corresponding R^2 values for the Cedar/Fir and Shrub PFTs.

553 Downgrading the spatial resolution of the AVIRIS imagery to 30 m and applying the
554 constrained MESMA method results in generally similar RMSEs and biases to constrained 18 m
555 MESMA: RMSEs increase by up to 4.2%, and biases change by -0.7% for Western Pines to 1%
556 for Western Hardwoods. Similarities in the 18 and 30 m Conventional MESMA are also shown
557 in the scatterplots of observed and predicted canopy composition (Figure A3), except for a
558 decrease in R^2 values for the Shrub PFT. The Bare-ground and Grass/NPV endmember classes in
559 the constrained 18 m MESMA had high RMSEs of 49% and 51% and high bias (26% and -28%
560 bias), but a significantly better combined Grass/NPV+ Bare-ground RMSE of 10.5% and bias of
561 -2.3% (30 m product RMSEs of 16.8% and bias of -5.1%). This represents a 24% and 17%
562 improvement in combined Grass/NPV+ Bare-ground RMSE and Bias of the 18 m constrained
563 versus conventional MESMA methods.

564 Figure 8 shows the fraction of wavelengths in the spring and fall that had JM distance
565 scores of 0.9 or above, for each pairwise combination of PFTs as well as each PFT and Bare-
566 ground. JM distance values asymptote at 2.0, indicating perfect separability (Richards and Jia,
567 2006). As the figure indicates, with the exception of Bare-ground /PFT pairs, levels of

separability in the fall were generally equal to or higher than in the spring. The overall level of spectral separability indicated that of the twenty-eight PFT/PFT plus PFT/ Bare-ground pairs, seventeen had high values of 0.9 or greater in more than half of the different band-seasons. The lowest levels of separability occurred for the Western Pine/High-elevation Pine combination (24% of spring wavelengths, 30% of fall wavelengths), Shrub/ Bare-ground (32% spring, 27% fall), Grass/NPV and Bare-ground (29% spring, 34% fall), and Western Hardwood/Western Pine (16% spring, 45% fall).

4. Discussion

Knowledge regarding the large-scale spatial distribution of canopy composition across heterogeneous landscapes is critical for improving terrestrial biosphere model predictions of carbon, water and energy fluxes, and how these will be affected by changes in climate, atmospheric CO_2, and land-use. Existing national and international large-scale ecosystem canopy composition products derived from multi-spectral imagery such as that shown in Figure 1a, do not have the fine-scale diversity and heterogeneity required by modern-day individual-based terrestrial biosphere models. In this study, we evaluated the ability of AVIRIS Airborne imaging spectrometry data to provide spatially-resolved estimates for the fractional composition of seven plant functional types (PFTs): Grasses/NPV, Shrubs, Oaks, Western Hardwoods, Western Pines, Cedar/Fir, and High-elevation Pines, across a functionally diverse and topographically heterogeneous ~710 km^2 area in the Southern Sierra Mountains of California via Multiple Endmember Spectral Mixture Analysis (MESMA).

Our results show that conventional MESMA, in which all seven putative PFT endmembers were included as potentially present in all pixels, captured the large-scale spatial

590 variation in the dominant PFT composition across the study area well (Figure 2c), but closer
591 examination revealed errors in the estimated spatial distributions and fractional abundance
592 estimates of the seven PFTs at the four flux tower sites (Figure 4). Specifically, conventional
593 MESMA incorrectly estimated that a) Grass/NPV, Shrub and Oak PFTs were present at low
594 abundances in high elevation areas; b) a widespread abundance of the Western Hardwood PFT,
595 which should have been restricted to mid- elevations; and c) a low abundance of the High-
596 elevation Pine PFT in low-elevation areas where it should have been absent (Figure 3). Although
597 high-elevation grasslands can exist, the large abundance estimated using conventional MESMA
598 (Figure 7) is not reflected in the CALVEG dataset (Figure 2b) or NLCD (Figure 1a).
599 Furthermore, RMSEs and biases resulting from comparisons of the fractional abundances of the
600 woody PFTs estimated with conventional MESMA against fractional abundances observed in
601 ground-based inventory plots (Figure 5 & Table 4), confirmed these inaccuracies.

602 Constrained MESMA, in which auxiliary information was used to limit the PFT
603 endmembers considered when classifying a given pixel, captured the observed large-scale spatial
604 variation in the dominant PFT composition across the study area without the problem of PFTs
605 being present at low-abundance in areas where geospatial data and ground inventories indicated
606 that they are absent (Figure 2b). Consequently, constrained MESMA resulted in modest
607 improvements of PFT fractional abundance estimates compared to the conventional MESMA
608 using plot-level validation dataset (see Figure 4&5 & Table 4). The largest remaining error in the
609 constrained MESMA was in the estimates of the High-elevation Pine and Cedar/Fir PFTs
610 abundances as seen in Figure 4 for P301 and Shorthair, and in Cedar/Fir plot-level validations in
611 Figure 5. This confusion is a result of the similarity of endmember spectra: where only 47-55%
612 of High-elevation Pine and Cedar/Fir endmember pairs had a Jeffries-Matusita distance greater

than 0.9 (Figure 8). Other low spectral separability of PFTs occurred between Western and High-elevation Pine (24-30%) and Western Hardwood and Western Pine (16-45%); however, confusion of these PFTs was minimized through the constrained MESMA method due to minimal co-occurrences of these PFTs.

The relatively low levels of spectral separability of Grass/NPV and Bare-ground, and Shrub and Bare-ground spectra explains why Bare-ground fractional cover increases following the exclusion of Grasses and Shrubs from higher elevation areas (see panels a and b of Figures 3 and 6, and grey bars in Figure 7). The low separability of Grass/NPV and Bare-ground (29-34% in Figure 8) is also indicative of their low individual class accuracies (51% and 49% RMSEs Constrained MESMA / 55.4% and 34.1% RMSEs Conventional MESMA). The two AVIRIS images, taken in June and November, were not acquired during the active growing period of the abundant low-elevation grasses (Goulden et al. 2012), which may have contributed to the low separability between Grass/NPV and Bare-ground. However, with the constrained method, limiting co-occurrence of grasses, shrubs and bare-ground, resulted in a higher combined Grass/NPV and Bare-ground accuracy (RMSEs of 10.5% versus 34.5 % for constrained vs conventional MESMA).

Use of the constrained MESMA method also greatly reduced computational time as fewer signatures need to be evaluated per pixel: conventional MESMA required 189,428 unique models to be tested for each pixel, consuming over 37 days of computational time (Intel Core i7 3.40GHz processor) while constrained MESMA evaluated an average of 4,210 models per pixel (highly mixed pixels required up to 16,644 models, and highly homogeneous pixels required as few as 43 models), and took just over 1 day of computational time. These computational times

635 would decrease in a parallel or cloud computing environment; however, the relative speed
636 difference between the two methods of classification would remain.

637 Results from this study, reporting overall RMSEs of 11.9% and 11.3% for the
638 conventional and constrained MESMA approaches, are comparable to accuracies reported in the
639 literature: for example, Roth et al. (2015) reported overall kappa values in MESMA
640 classification at Southern Sierra Nevada mixed conifer and Santa Barbara shrub/woodland
641 ecosystems of 0.44 and 0.84 respectively, and reported that confusion was highest among species
642 such as *Calocedrus decurrens*, *Pinus jeffreyi*, *Pinus lambertiana*, *Pinus sabiniana* and *Quercus*
643 *douglasii*. Roth et al. (2012) used MESMA over Santa Barbara, with resulting overall accuracies
644 of 80%, but with several accuracies below 50% (e.g. e.g. Shrubs and Western Hardwoods).
645 Hamada et al. (2011) used MESMA with Quickbird imagery over Californian sage scrub
646 communities, estimating fractional cover to an overall RMSE of 13% (9-17.5% for individual
647 PFTs). In chaparral shrublands of California, a MESMA analysis of AVIRIS imagery by
648 Dennison and Roberts (2003) estimated species level maps with accuracies of up to 89% (67-
649 100% for individual species).

650 The results of this study build upon Antonarakis et al.'s (2014) analysis of PFT
651 composition in three important ways. First, Antonarakis et al. (2014) estimated spatial variation
652 in the fractional composition of five tree PFTs across a ~4 km^2 area within a single temperate
653 forest biome/land cover class. In contrast, this study estimated spatial variation in the fractional
654 composition of seven PFTs (including herbaceous and shrub as well as tree PFTs) distributed
655 over a ~710 km^2 region spanning a broad range of biome types, ranging from tree-grass
656 savannahs to high elevation conifer-forests (Figure 2d). Second, in contrast to the relatively flat
657 and near-continuous closed-canopy forest landscape analyzed by Antonarakis et al. (2014), the

region investigated in this study included significant topography (Figure 1b) and a landscape in which tree canopies are relatively sparse. Even when imagery has been orthorectified and the effects of topography incorporated into the atmospheric correction (see Section 2.1), topographic distortions remain. In addition, the relatively sparse tree canopies present in the region give rise to significant shadowing effects and influence from bare-ground within the imagery that increases within-class spectral variability (Fassnacht et al. 2016), which in turn, makes spectral un-mixing considerably more difficult than situations with near-continuous canopy cover (Asner et al., 2007, Asner et al., 2015).

Third, planned satellite-based imaging spectrometry platforms will produce measurements similar to those used in this study. The latest proposed configuration of the HyspIRI mission has a spectral resolution of 10 nm with bands spanning wavelengths from 0.38 to 2.5 μm, and at a spatial resolution of 30 m at nadir, comparable to Landsat (Lee et al., 2015). As shown in this study, applying the constrained MESMA to 30 m downscaled AVIRIS resulted in similar PFT spatial patterns (Figures 2e and 7) and abundance accuracies (Figures 4 and 5) albeit with individual PFT RMSEs up to 4.2% higher than the 18 m constrained method, and with < 1% change in bias for all woody PFTs. These results strongly demonstrate that measurements from HyspIRI, or another similar instrument, will be capable of providing important and much-needed information about the diversity and heterogeneity of plant canopy composition suitable for constraining terrestrial biosphere model predictions of large-scale ecosystem function.

Relevance to Terrestrial Ecosystem and Biosphere models: Remote-sensing Constrained Predictions of Terrestrial Carbon Fluxes

681 As noted in the Introduction, an important application of spatially-extensive estimates of
682 fractional PFT composition such as ones developed in this study (Figure 2, panels c-e) is to
683 constrain individual-based terrestrial biosphere model predictions of large-scale ecosystem
684 dynamics and function. An illustrative example of this is shown in Figure 9. Panel (a) shows
685 regional carbon fluxes for the study area predicted by the ED2 individual-based terrestrial
686 biosphere model (Medvigy et al. 2009) following ingestion of the 18-m constrained MESMA
687 estimate of ecosystem composition (Figure 2d) into a regional implementation of the model on a
688 1 km grid of climate forcing and soil properties. As the figure illustrates, Gross Primary
689 Productivity (GPP) Net Ecosystem Productivity (NEP) are lowest in the western, low-elevation
690 portion of the study area (< 1000m) that is comprised mostly of oak savanna ecosystems, highest
691 in the central, mid-elevation region (1000-2250 m) that is comprised of and pine/oak forests and,
692 at higher elevations within this range, montane and mixed conifer forests, and intermediate in the
693 eastern, high-elevation region (>2250m) that is largely comprised of sub-alpine pine forests
694 (compare Figure 9 bottom panel to Figure 2d).
695 Panel (b) shows predicted versus observed seasonality in GPP and NEP at Soaproot. As
696 the figure illustrates, when the ED2 model is initialized with HyspIRI-based (Constrained
697 MESMA) estimates of canopy composition (Figure 9 blue lines) show close agreement in
698 predicted GPP and NEP with initializations using ground-based estimates of canopy composition
699 (red lines) and observed fluxes (black lines and points), and large improvements over carbon
700 fluxes estimated from an equilibrium potential vegetation simulation (grey dashed lines). Further
701 details on this regional model implementation and further site-level evaluations can be found in
702 Antonarakis et al. (in prep).
703

Using Geospatial Data for Endmember Reduction: Challenges and Trade-Offs

As shown here, auxiliary geospatial data, such as CALVEG that provide information on the presence and absence of species across a region, can be used to constrain and improve MESMA-based classifications of canopy composition across large ecosystem gradients where the potential for ecologically-distinct, but spectrally-similar species or PFTs increases. There was a modest improvement in overall canopy composition accuracy from the conventional to constrained MESMA driven mainly by the large reduction in Western Hardwood RMSE, and constrained MESMA also limited the erroneous occurrence of PFTs being present at low abundances in areas outside their expected ecological ranges; specifically, the occurrence of High Pine and Cedar/Fir in low-elevation regions and the occurrence of Oak and Western Hardwoods at high elevations. Overall, constrained MESMA reduced the biases in PFT fractional abundances by 2-7.8%, and also significantly decreased computational time.

The endmember reduction approach used here is similar to that of Roth et al (2015), although here it is applied to a continuous scene rather than several distinct regions. Other studies have used different endmember reduction/pruning approaches. For example, Degerickx et al. (2017) reduced the endmembers considered for each pixel by calculating the distance of each library spectrum to eigenvectors from subset images and applying a spectral distance rejection threshold. Other studies (e.g. Garcia-Haro et al. 2005; Liu and Yang 2013) have used a-priori classifications or land cover masks (e.g. masking urban vs non-urban environments) to reduce the number of endmembers locally.

In all cases, the goal of endmember reduction methods is to reduce the occurrence of locally irrelevant classes from imagery collected over complex and heterogeneous environments. Use of available geospatial data to identify putative PFT endmembers used by Roth et al. (2015)

and this study will likely be essential in the context of a global-scale mission where the number of potential PFT combinations will overwhelm a conventional MESMA approaches to estimating spatial variation in PFT abundances. This will be particularly true for modern terrestrial biosphere and ecosystem models that incorporate sub-grid scale information on plant canopy composition that is used to define the vegetation attributes of a heterogeneous and functionally diverse plant canopy (e.g. Verheijen et al. 2013; Fyllas et al. 2014; Sakschewski et al. 2016; Levine et al. 2016).

Producing accurate estimates of PFT composition from future global imaging spectrometry data will, however, inevitably require an appropriate trade-off between use of geospatial data for endmember selection while allowing for uncertainty and error in available geospatial information regarding the presence or absence of species or plant functional type in those locations. For example, in their assessment of the CALVEG dataset used in this study, Milliken et al. (1998), showed that the accuracy of geospatial data varied across regions and across plant types. Accounting for uncertainty and error in geospatial data such as CALVEG is particularly important given the ongoing changes in spatial distributions of plant species as a result of climate change, species introductions, and other human activities. In the context of future global-scale missions, currently available geospatial datasets that could be used to constrain the endmember spectra of tree species considered for a given pixel could include the European Atlas of Forest Tree Species (San-Miguel-Ayanz et al. 2016), the tree species distribution maps available via the European Data Portal (https://www.europeandataportal.eu/), and the US Forest Service's Nationwide Datasets of Tree Species Distributions (Wilson et al. 2014). However, comparable datasets providing information on the presence and absence of shrubs and herbaceous plants will also be required.

Future Work:

Several avenues exist for further improving imaging spectrometry-derived estimates of canopy composition. First, composition estimates will be improved by the development of detailed spectral libraries for different PFTs. PFTs are usually comprised of multiple species; however, in cases where a PFT is comprised of many species with diverse leaf properties, it will tend to exhibit a significant degree of within-PFT spectral diversity thereby reducing its spectral separability from other PFTs. One approach to this challenge is to explicitly incorporate spectral similarity and/or differences in spectrally-separable plant canopy attributes into future definitions of the PFTs distinguished in terrestrial ecosystem models. Several studies (e.g. Gillon et al. 2009; Asner and Martin 2009, 2011, 2016; Martin et al., 2018) have concluded that it is possible to measure a number of leaf biochemical attributes such as nitrogen, phosphorous and lignin content. However, the mechanistic basis for observed correlations between reflectance spectra and these different biochemical attributes of leaves has been questioned by others (e.g. Knyazikhin et al. 2013a, 2013b; Wang et al. 2017) due the confounding effects of correlated variations in leaf and canopy structural attributes that also significantly affect reflectance spectra (see Ustin et al. 2013 for a review of this topic). However, there appears to greater consensus that other important leaf attributes, in particular, foliar chlorophyll content and Specific Leaf Area (SLA; or its reciprocal Leaf Mass per unit Area, LMA) are more directly measured with imaging spectrometry (Fourty et al. 1996; Baret and Fourty 1997; Asner and Martin 2009, 2011, 2016; Martin et al., 2018).

Second, in situations where identifying endmembers from satellite-acquired pixels (nominally ~30 m resolution) was difficult due to sparse canopy cover, high-resolution airborne

imagery could improve classification accuracy by a) assessing whether a pixel is composed of a contiguous target canopy (as done in this study), and b) extracting pure endmembers providing the airborne imagery is concurrent with and radiometrically calibrated to the satellite imagery (see McCorkel et al., 2016). Third, as seen in Figures 3 and 6, Bidirectional Reflectance Distribution Function (BRDF) effects are likely to have affected the accuracy achieved in this study. BRDF correction would likely improve the accuracy and reduce biases in the fractional cover estimates. Note however, the magnitude of BRDF effects are likely be smaller in imagery acquired from satellite-based platform: the latest HyspIRI mission configuration has a field-of-view of around 16^{o} (Lee et al., 2015), compared to imagery used in this study that had a field- of-view of approximately 35^{o}.

5. Conclusions

This study evaluated the ability of imaging spectrometry measurements, collected over a functionally and topographically heterogeneous ~710 km^2 area in the Southern Sierra Mountains of California, to estimate sub-pixel fractions of seven PFTs; Grass/NPV, Shrub, Oak, Western Hardwood, Western Pine, Cedar/Fir, and High-elevation Pine for use in terrestrial biosphere and ecosystem model simulations. Applying the conventional MESMA approach, in which all PFTs are evaluated for every pixel, to the original 18 m resolution dataset captured the major elevation-related shifts in canopy composition that occurred within the study area. Using a constrained MESMA approach, where the putative PFTs within each pixel were constrained using available geospatial data, improved the overall accuracy and reduced the large biases in the composition estimates of Western Hardwoods and Oaks PFTs. Downscaling the remote measurements to 30 m resolution, the planned spatial resolution of the proposed global-scale HyspIRI imaging spectrometer, resulted in a small reduction in accuracy of the constrained

method, but with minimal change in bias. The results of this study demonstrate that imaging spectrometry measurements from planned satellite missions such as HyspIRI, EnMAP, and HISUI will provide important and much-needed information about fine-scale heterogeneity in canopy composition necessary for improving terrestrial biosphere model simulations of regional- and global-scale vegetation dynamics and function.

Acknowledgements

The authors would like to acknowledge the NASA HyspIRI Preparatory Activity NNH11ZDA001N-HYSPIRI grant "Linking Terrestrial Biosphere Models with Remote Sensing Measurements of Ecosystem Composition, Structure, and Function". We thank Michael Goulden and Anne Kelly at the University of California, Irvine for the forest inventory plot data at the Southern Sierra CZO flux tower sites. We thank Susan Ustin and Dr. Huesca-Martinez at the University of California, Davis and the CZO/NEON projects for their ground-based inventory plots data across Southern Sierra CZO study region. We thank Robert Green, Ian McCubbin, Simon Hook, David Thompson and the HyspIRI Airborne Campaign Team for the AVIRIS flights, data, and pre-processing of images. We thank Dar Roberts for his guidance and assistance on the use of ViperTools 2.0. Development of ViperTools 2.0 was funded by the Belgian Science Policy Office in the framework of the STEREO III Programme Project VIPER2 (SR/xx/171), and NASA grant #NNX12AP08G).

References

Adams, J. B., Smith, M. O., & Johnson, P. E. (1986). Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander 1 site. *Journal of Geophysical Research: Solid Earth*, **91**(B8), 8098-8112.

Adams, J.B., and Gillespie, A.R. (2006). Chapter 4 Spectral Mixture Analysis. In Remote Sensing of Landscapes with Spectral Images: *A Physical Modeling Approach*, (Cambridge University Press), p.126-165.

Antonarakis, A.S., Bogan, S., Goulden, M.L., Moorcroft, P.R. (in prep). Impacts of the 2012-2015 Californian Drought on Carbon, Water and Energy Fluxes in Californian Sierras: Results from an Imaging Spectrometry-Constrained Terrestrial Biosphere Model.

Antonarakis, A.S., Saatchi, S.S., Chazdon, R.L., Moorcroft, P.R. (2011). Using Lidar and Radar measurements to constrain predictions of forest ecosystem structure and function. *Ecological Applications*, **21**, 1120-1137.

Antonarakis, A.S., Munger, J.W., and Moorcroft, P.R. (2014). Imaging spectroscopy- and lidar-derived estimates of canopy composition and structure to improve predictions of forest carbon fluxes and ecosystem dynamics: *Geophysical Research Letters* **41**, 2535–2542.

Asner, G.P., Knapp, D.E., Kennedy-Bowdoin, T., et al. (2007). Carnegie Airborne Observatory: in-flight fusion of hyperspectral imaging and waveform light detection and ranging (Lidar) for three-dimensional studies of ecosystems. *Journal of Applied Remote Sensing*, **1**, 013536.

Asner, G.P., Knapp, D.E., Kennedy-Bowdoin, T., Jones, M.O., Martin, R.E., Boardman, J., and Hughes, R.F. (2008). Invasive species detection in Hawaiian rainforests using airborne imaging spectroscopy and Lidar. *Remote Sensing of Environment*, **112**, 1942–1955.

Asner, G.P. and R.E. Martin. (2009). Airborne spectranomics: Mapping canopy chemical and taxonomic diversity in tropical forests. *Frontiers in Ecology and the Environment*, **7**, 269-276.

Asner, G.P., and R.E. Martin. (2011). Canopy phylogenetic, chemical and spectral assembly in a lowland Amazon forest. *New Phytologist*, **189**, 999-1012.

Asner, G.P., Martin, R.E., Anderson, C.B. and Knapp, D.E., (2015). Quantifying forest canopy traits: Imaging spectroscopy versus field survey. *Remote Sensing of Environment*, **158**, 15-27.

Asner, G.P. and Martin, R.E., (2016). Spectranomics: Emerging science and conservation opportunities at the interface of biodiversity and remote sensing. *Global Ecology and Conservation*, **8**, 212-219.

Atkin, O.K., Bloomfield, K.J., Reich, P.B., Tjoelker, M.G., Asner, G.P., Bonal, D. et al. (2015). Global variability in leaf respiration in relation to climate, plant functional types and leaf traits. New Phytol., **206**, 614–636.

Baccini, A. et al. (2012). Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature Climate Change*, **2**, 182-185.

Baret, F. and Fourty, T. (1997). Radiometric estimates of nitrogen status of leaves and canopies. In Diagnosis of the nitrogen status in crops, pages 201-227. Springer, Berlin.

Bontemps, S., Defourny, P., Bogaert, E.V., Arino, O., Kalogirou, V. and Perez, J.R., (2011). *GLOBCOVER 2009-Products description and validation report*. Vancouver.

Cade, B.S. 1997. Comparison of Tree Basal Area and Canopy Cover in Habitat Models: Subalpine Forest. *Journal of Wildlife Management*, **61**, 326–35.

CALVEG - Existing Vegetation, [ESRI geodatabase]. (2009a). McClellan, CA: USDA-Forest Service, Pacific Southwest Region. ExistingVegCentralValley1998_2007_v1. [Accessed 2014].

CALVEG - Existing Vegetation, [ESRI geodatabase]. (2009b). McClellan, CA: USDA-Forest Service, Pacific Southwest Region. ExistingVegGreatBasin1999_2009_v1. [Accessed 2014].

CALVEG - Existing Vegetation, [ESRI geodatabase]. (2009c). McClellan, CA: USDA-Forest Service, Pacific Southwest Region. ExistingVegSouthSierra2000_2008_v1. [Accessed 2014].

CALVEG - Existing Vegetation, [ESRI geodatabase]. (2010). McClellan, CA: USDA-Forest Service, Pacific Southwest Region. ExistingVegNorSierra2000_2009_v1. [Accessed 2014].

Clark, M.L. (2017). Comparison of simulated hyperspectral HyspIRI and multispectral Landsat 8 and Sentinel-2 imagery for multi-seasonal, regional land-cover mapping. *Remote Sensing of Environment*, **200**, 311-325.

Clark, M.L., Buck-Diaz, J. and Evens, J., (2018). Mapping of forest alliances with simulated multi-seasonal hyperspectral satellite imagery. *Remote Sensing of Environment*, **210**, 490-507.

Degerickx, J., Okujeni, A., Iordache, M.D., Hermy, M., van der Linden, S. and Somers, B., (2017). A novel spectral library pruning technique for spectral unmixing of urban land cover. *Remote Sensing*, **9**(6), p.565.

Dennison, P.E. and Roberts, D.A., (2003). Endmember selection for multiple endmember spectral mixture analysis using endmember average RMSE. *Remote Sensing of Environment*, **87**(2-3), pp.123-135.

Drake, J.B., Dubayah, R.O., Knox, R.G., Clark, D.B. and Blair, J.B., 2002. Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest. *Remote Sensing of Environment*, **81**, 378-392.

Dubayah, R.O., Drake, J.B. (2000). Lidar remote sensing for forestry. *Journal of Forestry*, **98**, 44-46.

Fassnacht, F.E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L.T., Straub, C., Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed

data. *Remote Sensing of Environment*, **186**, 64–87. http://dx.doi.org/10.1016/j.rse.2016.08.013

Féret, J.B. and Asner, G.P., (2013). Tree species discrimination in tropical forests using airborne imaging spectroscopy. *IEEE Transactions on Geoscience and Remote Sensing*, **51**(1), 73-84.

Ferreira, M.P., Zortea, M., Zanotta, D.C., Shimabukuro, Y.E. and de Souza Filho, C.R., (2016). Mapping tree species in tropical seasonal semi-deciduous forests with hyperspectral and multispectral data. *Remote Sensing of Environment*, **179**, 66-78.

Fisher, J.B., Huntzinger, D.N., Schwalm, C.R., and Sitch, S. (2014). Modeling the Terrestrial Biosphere. *Annual Review of Environment and Resources*, **39**, 91–123.

Fourty, T., Baret, F., Jacquemoud, S., Schmuck, G., & Verdebout, J. (1996). Leaf optical properties with explicit description of its biochemical composition: Direct and inverse problems. *Remote Sensing of Environment*, **56**(2), 104-117.

Friedl, M.A. et al (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets, *Remote Sensing of Environment*, **114**, 168-182.

Fyllas, N.M., Gloor, E., Mercado, L.M., Sitch, S., Quesada, C.A., Domingues, T.F., Galbraith, D.R., Torre-Lezama, A., Vilanova, E., Ramírez-Angulo, H. and Higuchi, N., (2014). Analysing Amazonian forest productivity using a new individual and trait-based model (TFS v. 1). *Geoscientific Model Development*, **7**(4), 1251-1269.

García-Haro, F.J., Sommer, S., Kemper, T. (2005). A new tool for variable multiple endmember spectral mixture analysis (VMESMA). *Int. J. Remote Sens.*, **26**, 2135–2162.

Gillon, D., Houssard, C., & Joffre, R. (1999). Using near-infrared reflectance spectroscopy to predict carbon, nitrogen and phosphorus content in heterogeneous plant material. *Oecologia*, **118**(2), 173-182.

Goodenough, D.G., Dyk, A., Niemann, K.O., Pearlman, J.S., Chen H., Han, T., Murdoch, M., and West, C. (2003). Processing Hyperion and ALI for forest classification. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, 1321–1331.

Goulden, M.L., Anderson, R.G., Bales, R.C., Kelly, A.E., Meadows, M. and Winston, G.C., (2012). Evapotranspiration along an elevation gradient in California's Sierra Nevada. Journal of Geophysical Research: *Biogeosciences*, **117**(G3).

Green, R.O., Eastwood, M.L., Sarture, C.M., Chrien, T.G., Aronsson, M., Chippendale, B.J., Faust, J.A., Pavri, B.E., Chovit, C.J., Solis, M. and Olah, M.R., 1998. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). *Remote sensing of environment*, **65**(3), pp.227-248.

Guanter, L., Kaufmann, H., Forster, S. et al. (2016) EnMAP Science Plan. EnMAP Technical Report, Potsdam GFZ Data Services. 69 p. doi: 10.2312/enmap.2016.006

Hamada, Y., Stow, D.A., and Roberts, D.A. (2011). Estimating life-form cover fractions in California sage scrub communities using multispectral remote sensing. *Remote Sensing of Environment*, **115**, 3056–3068.

Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., and Megown, K., (2015), Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing*, **81**, 345-354.

Hochberg, E.J., Roberts, D.A., Dennison, P.E. and Hulley, G.C. (2015). Special issue on the Hyperspectral Infrared Imager (HyspIRI): Emerging science in terrestrial and aquatic ecology, radiation balance and hazards. *Remote Sensing of Environment*, **167**, 1-5.

Huntzinger, D.N., Post, W.M., Wei, Y., Michalak, A.M., West, T.O., Jacobson A.R., Baker I.T., Chen, J.M., Davis, K.J., Hayes, D.J., Hoffman, F.M., Jain, A.K., Liu, S., McGuire, A.D., Neilson, R.P., Potter, C., Poulter, B., Price, D., Raczka, B.M., Tian, H.Q., Thornton, P., Tomelleri, E., Viovy, N., Xiao, J., Yuan, W., Zeng, N., Zhao, M., Cook, R. North American Carbon Program (NACP) regional interim synthesis: terrestrial biospheric model intercomparison, *Ecological Modelling*, **232**, 144–15.

Hurtt, G.C., Dubayah, R., Drake, J., Moorcroft, P.R., Pacala, S.W., Blair, J.B., and Fearon, M.G. (2004). Beyond potential vegetation: combining lidar data and a height-structured model for carbon studies. *Ecological Applications*, **14**, 873–883.

Hurtt, G.C. et al. (2010) Linking Models and Data on Vegetation Structure: Quantifying Model-Data Requirements for Future Space-borne Missions, *Journal of Geophysical Research*, **115** G00E10.

Kattge, J., Diaz, S., Lavorel, S., Prentice, I.C., Leadley, P., Bönisch, G., Garnier, E., Westoby, M., Reich, P.B., Wright, I.J. and Cornelissen, J.H.C., (2011). TRY–a global database of plant traits. *Global Change Biology*, **17**(9), 2905-2935.

Kelly, A.E. (2014). *Climate controls on ecosystem production, biomass, and water cycling*. Dissertation. University of California, Irvine.

Keshava, N., and Mustard, J.F. (2002). Spectral unmixing. *IEEE Signal Processing Magazine*, **19**, 44–57.

Knyazikhin, Y., Schull, M.A., Stenberg, P., Mõttus, M., Rautiainen, M., Yang, Y., Marshak, A., Carmona, P.L., Kaufmann, R.K., Lewis, P. and Disney, M.I. (2013a). Hyperspectral remote sensing of foliar nitrogen content. *Proceedings of the National Academy of Sciences*, **110**(3), E185-E192.

Knyazikhin, Y., Lewis, P., Disney, M.I., Stenberg, P., Mottus, M., Rautianinen, M., Kaufmann, R.K., Marshak, A., Schull, M.A., Carmona, P.L. and Vanderbilt, V. (2013b). Decoupling Contributions from Canopy Structure and Leaf Optics is Critical for Remote Sensing Leaf Biochemistry (Reply to Townsend, et al.).

Kokaly, R.F., Despain, D.G., Clark, R.N., and Livo, K.E. (2003). Mapping vegetation in Yellowstone National Park using spectral feature analysis of AVIRIS data. *Remote Sensing of Environment*, **84**, 437–456.

Lee, C.M., Cable, M.L., Hook, S.J., Green, R.O., Ustin, S.L., Mandl, D.J. and Middleton, E.M. (2015). An introduction to the NASA Hyperspectral InfraRed Imager (HyspIRI) mission and preparatory activities. *Remote Sensing of Environment*, **167**, 6-19.

Lefsky, M.A., Harding, D.J., Keller, M., Cohen, W.B., Carabajal, C.C., Del Bom Espirito-Santo, F., Hunter, M.O. and de Oliveira, R. (2005). Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical research letters*, **32**(22).

Levine, N.M. et al. (2016). Ecosystem heterogeneity determines the ecological resilience of the Amazon to climate change. *Proceedings of the National Academy of Sciences*,**113** (3), 793-797.

Li, L., Ustin, S.L., and Lay, M. (2005). Application of multiple endmember spectral mixture analysis (MESMA) to AVIRIS imagery for coastal salt marsh mapping: A case study in China Camp, CA, USA. *International Journal of Remote Sensing*, **26**, 5193–5207.

Liu, T. and Yang, X., 2013. Mapping vegetation in an urban area with stratified classification and multiple endmember spectral mixture analysis. *Remote Sensing of Environment*, 133, 251-264.

McCorkel, J., Cairns, B. and Wasilewski, A., (2016). Imager-to-radiometer in-flight cross calibration: RSP radiometric comparison with airborne and satellite sensors. *Atmospheric Measurement Techniques*, **9**(3), 955-962.

Martin, M.E., Newman, S.D., Aber, J.D., and Congalton, R.G. (1998). Determining forest species composition using high spectral resolution remote sensing data. *Remote Sensing of Environment*, **65**, 249–254.

Martin, R.E., Chadwick, K.D., Brodrick, P.G., Carranza-Jimenez, L., Vaughn, N.R. and Asner, G.P. (2018). An Approach for Foliar Trait Retrieval from Airborne Imaging Spectroscopy of Tropical Forests. *Remote Sensing*, **10**(2), 199.

Matsunaga, T., Iwasaki, A., Tsuchida, S., Iwao, K., Tanii, J., Kashimura, O., Nakamura, R., Yamamoto, H., Kato, S., Mouri, K. and Tachikawa, T., 2016, July. Current status of Hyperspectral Imager Suite (HISUI) and its deployment plan on International Space Station. *In Geoscience and Remote Sensing Symposium (IGARSS)*, 2016 IEEE International (pp. 257-260). IEEE.

Milliken, J., Beardsley, D., Gill, S. and Warbington, R., (1998). Accuracy assessment of a vegetation map of northeastern California using permanent plots and fuzzy sets. In Natural resource management using remote sensing and GIS: proceedings of the Seventh Forest Service Remote Sensing Applications Conference, Nassau Bay, Texas, April 6-10, 1998/Jerry Dean Greer, editor. Bethesda, Md.: *American Society for Photogrammetry and Remote Sensing*, c1998.

Moorcroft, P.R (2006) (Invited feature, 20th anniversary issue), How close are we to a predictive science of the biosphere? *Trends in Ecology and Evolution*, 21: 400-407.

National Elevation Dataset (NED), [Raster Data]. (2009). Sioux Falls, SD: U.S. Geological Survey (USGS). [Accessed 2015].

National Academies of Sciences, Engineering, and Medicine, (2018). Thriving on Our Changing Planet: A Decadal Strategy for Earth Observation from Space. Washington, DC: *The National Academies Press.* https://doi.org.10.17226/24938, p.1-716.

Quaife, T., Quegan, S., Disney, M., Lewis, P., Lomas, M. and Woodward, F.I., (2008). Impact of land cover uncertainties on estimates of biospheric carbon fluxes. *Global Biogeochemical Cycles*, **22**, 1-12.

Richards, J.A., and Jia, X. (2006). *Remote sensing digital image analysis: an introduction* (Berlin: Springer).

Roberts, D., Halligan, K., Dennison, P. (2007). VIPER Tools User Manual Version 1.5 https://www.scribd.com/document/141164870/VIPER-Tools-User-Manual-v1-5.

Roberts, D.A., Gardner, M., Church, R., Ustin, S., Scheer, G., and Green, R.O. (1998). Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. *Remote Sensing of Environment*, **65**, 267–279.

Roth, K.L., Dennison, P.E. and Roberts, D.A., (2012). Comparing endmember selection techniques for accurate mapping of plant species and land cover using imaging spectrometer data. *Remote Sensing of Environment*, **127**, pp.139-152.

Roth, K.L., Roberts, D.A., Dennison, P.E., Alonzo, M., Peterson, S.H., and Beland, M. (2015). Differentiating plant species within and across diverse ecosystems with imaging spectroscopy. *Remote Sensing of Environment*, **167**, 135–151.

Saatchi, S.S., Houghton, R.A., Dos Santos Alvala, R.C., Soares, J.V. and Yu., Y. (2007). Distribution of aboveground live biomass in the Amazon. *Global Change Biology*, **13**, 816-837.

Saatchi, S.S., Harris, N. L., Brown, S. et al. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences*, **108**, 9899-9904.

Sabol, D.E. (1992). Quantitative Subpixel Spectral Detection of Targets in Multispectral Images. *Journal of Geophysical Research*, **97,** 2659–2672.

Sakschewski, B., von Bloh, W., Boit, A., Poorter, L., Peña-Claros, M., Heinke, J., Joshi, J. and Thonicke, K., (2016). Resilience of Amazon forests emerges from plant trait diversity. *Nature Climate Change*, http://dx.doi.org/10.1038/nclimate3109, 1-5.

Thompson, D.R., Gao, B.-C., Green, R.O., Roberts, D.A., Dennison, P.E., and Lundeen, S.R. (2015). Atmospheric correction for global mapping spectroscopy: ATREM advances for the HyspIRI preparatory campaign. *Remote Sensing of Environment*, **167**, 64–77.

Ustin, S.L., 2013. Remote sensing of canopy chemistry. *Proceedings of the National Academy of Sciences*, **110**(3), 804-805.

van Aardt, J.A.N., and Wynne, R.H. (2007). Examining pine spectral separability using hyperspectral data from an airborne sensor: An extension of field-based results. *International Journal of Remote Sensing*, **28,** 431–436.

Verheijen, L.M., Brövkin, V., Aerts, R., Bonisch, G., Cornelissen, J.H., Kattge, J., Reich, P.B., Wright, I.J. and Van Bodegom, P.M., (2013). Impacts of trait variation through observed trait-climate relationships on performance of an Earth system model: a conceptual analysis. *Biogeosciences*, **10**, 5497–5515.

Wang, Z., Skidmore, A.K., Wang, T., Darvishzadeh, R., Heiden, U., Heurich, M., Latifi, H. and Hearne, J., (2017). Canopy foliar nitrogen retrieved from airborne hyperspectral imagery by correcting for canopy structure effects. *International journal of applied earth observation and geoinformation*, **54**, 84-94.

Wilson, B.T., Riemann, R. (2014). *USDA Forest Service Nationwide Datasets of Tree Species Distributions Created*. Northern Research Station, Accessed online (https://www.fs.fed.us/research/highlights/highlights_display.php?in_high_id=630).

Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J. H. C., Diemer, M., Flexas, J., Garnier, E., Groom, P. K., Gu- lias, J., Hikosaka, K., Lamont, B. B., Lee, T., Lee, W., Lusk, C., Midgley, J. J., Navas, M. L., Ni-inemets, U., Oleksyn, J., Osada, N., Poorter, H., Poot, P., Prior, L., Pyankov, V. I., Roumet, C., Thomas, S. C., Tjoelker, M. G., Veneklaas, E. J., and Villar, R., (2004). The worldwide leaf economics spectrum, Nature, **428**, 821–827.

Youngentob, K.N., Roberts, D.A., Held, A.A., Dennison, P.E., Jia, X. and Lindenmayer, D.B., (2011). Mapping two Eucalyptus subgenera using multiple endmember spectral mixture analysis and continuum-removed imaging spectrometry data. *Remote Sensing of Environment*, **115**(5), 1115-1128.

Figure Captions

Figure 1. (a) Land cover classification from the 2011 National Land Cover Database (NLCD). b) Plant associations map from USDA Forest Service 2009 CALVEG database. (b) False color composite of AVIRIS imaging spectrometer data collected in the Sierra Nevada, northwest of Fresno, California in November 2013 for the ~710 km^2 area analyzed in this study. Yellow stars show the locations of the four eddy-flux towers are within the study area. (c) Elevation map of the study area showing the east-to-west elevation gradient across the region. (d) Red panel shows location of the study area within California.

Figure 2. (a) Land cover classification from the 2011 National Land Cover Database (NLCD). b) Plant associations map from USDA Forest Service 2009 CALVEG database. Legend indicates the Plant Functional Types (Table 2) associated with each formation. PFT names are abbreviated as follows: CF = Cedar/Fir, GR = Grass, HP = High-elevation Pine, OK = Oaks, SH = Shrub, WH = Western Hardwood, WP = Western Pine. For clarity purposes, only the eighteen most present Plant Associations are shown in the legend. c) Dominant PFT fraction as classified by the conventional MESMA method. (d) Dominant PFT fraction classified by the constrained MESMA method using 18 m and (e) downscaled 30 m AVIRIS data. The grey dashed box in the top left of panel (d) shows, for comparative purposes, the size of the region analyzed by Antonarakis et al (2014).

Figure 3. MESMA fractional PFT estimates derived from 18 m AVIRIS reflectance using the conventional MESMA method (every PFT combination tested on each pixel). Cold colors represent low fractions, and warm colors represent high fractions on a scale from zero to one.

Figure 4. Fractional abundance of the tree PFTs at each of the four eddy flux tower sites and the abundance estimates from the conventional and constrained MESMA methods. The locations of the flux towers are shown in Figure 1b.

Figure 5. Plot-level Validation of the conventional MESMA at 18 m (solid bars), and the Constrained MESMA at 18 m (shaded bars) and downscaled 30 m (horizontal striped bars) using ninety-five forest inventory plots. MESMA classifications are compared to the horizontal crown

area occupied by each PFT in a field plot. Note: none of the ninety-five plots contained the High Pine PFT so non-zero estimates for this PFT indicate commission errors.

Figure 6. Proportional MESMA fractional PFT estimates derived from 18 m AVIRIS reflectance using the constrained MESMA method. Cold colors represent low fractions, and warm colors represent high fractions on a scale from zero to one.

Figure 7. Proportional distribution of plant functional types (PFTs) as a function of elevation estimated by the conventional MESMA method and the constrained MESMA method using 18 m and downscaled 30 m AVIRIS data.

Figure 8. Percent of wavelengths in each endmember pair that scored above 0.9 JM distance for signatures taken from the spring (June) image (dark blue bars) and signatures taken from the fall (November) image (light blue bars).

Figure 9. Regional and site-level carbon flux predictions obtained from the Ecosystem Demography (ED2) model using the regional estimate of fractional PFT composition produced in this study. The model was implemented at 1km resolution using meteorological forcing specified from the North American Land Data Assimilation System (NLDAS-2) and soil properties specified from the United States Geological Survey Soil Survey Geographic Database (SSURGO) of soil texture and depth. Ecosystem composition was specified using the 18-m constrained MESMA (Figure 2d). Panel (a) shows the predicted spatial patterns of Gross Primary Productivity (GPP) and Net Ecosystem Productivity (NEP) in 2010. The star symbol in the GPP panel indicates the location of the Soaproot flux tower site. The lower panel (b) shows predicted versus observed seasonal carbon fluxes (GPP and NEP) at the Soaproot CZO flux-tower site. The black lines and points are the observed carbon fluxes; the red lines are the predictions of the ED2 model when initialized with ground-based estimates of canopy composition; the blue-lines are the predictions of the ED2 model using HyspIRI-based estimate of canopy composition plotted in Figure 2d; and the grey dashed lines show the ED2 predictions from a potential vegetation simulation. The canopy composition of each simulation is shown on the right of panel b.

1123

Figure 1

Click here to download high resolution image

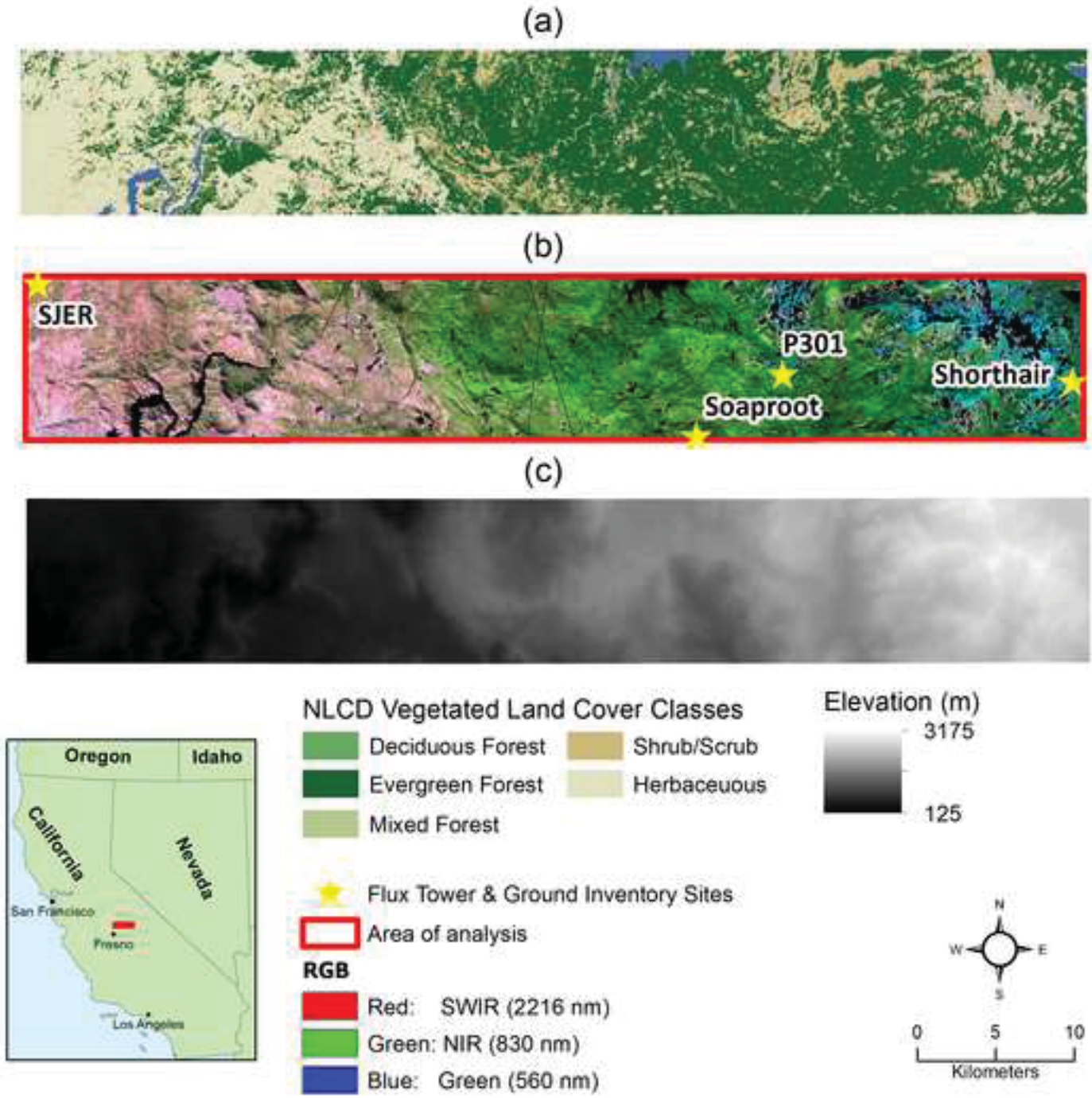
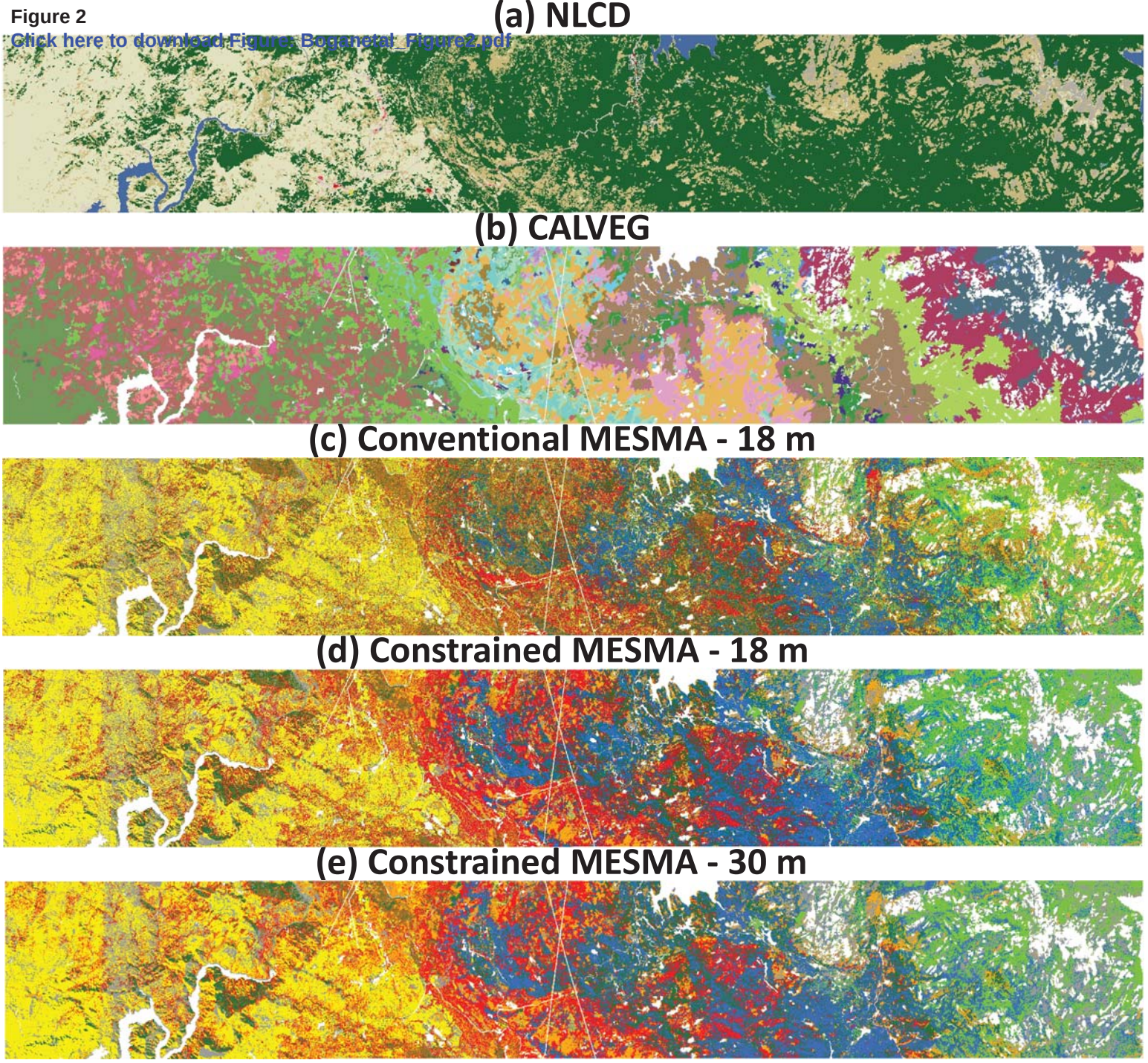


Figure 2



NLCD Vegetated Land Cover Classes

- Deciduous Forest
- Evergreen Forest
- Mixed Forest
- Shrub/Scrub
- Herbaceuous

N
W E
S
0 5 10
Kilometers

MESMA PFTs

- BARE-GROUND
- CEDARFIR
- GRASS/NPV
- HIGHPINE
- OAK
- SHRUB
- WESTERNHDWD
- WESTERNPINE

CALVEG Associations

- Blue Oak [GR, OK, SH, WP]
- Annual Grasses and Forbs [GR, WH, WP]
- Mixed Conifer - Pine [CF, OK, SH, WP]
- Ponderosa Pine & Canyon Live Oak [CF, OK, SH, WP]
- Interior Live Oak [OK, SH, WH, WP]
- Red Fir [CF, HP]
- Subalpine Conifers [CF, HP]
- Canyon Live Oak [OK, SH]
- Ponderosa Pine [CF, OK, SH, WP]
- Lower Montane Mixed Chaparral [SH, WH]
- Gray Pine & Interior Live Oak [GR, OK, SH, WP]
- Mixed Conifer - Fir [CF, HP, SH, WP]
- Gray Pine & Blue Oak [GR, OK, SH, WP]
- Mixed Conifer - Pine & Black Oak [CF, GR, HP, OK, SH, WP]
- Ponderosa Pine & Black Oak [CF, HP, PK, SH, WP]
- Mixed Conifer - Pine & Canyon Live Oak [CF, OK, SH, WP]
- Upper Montane Mixed Chaparral [SH]
- Black Oak [CF, HP, OK, SH, WP]

Figure 3
Click here to download Figure: Boganetal_Figure3.pdf

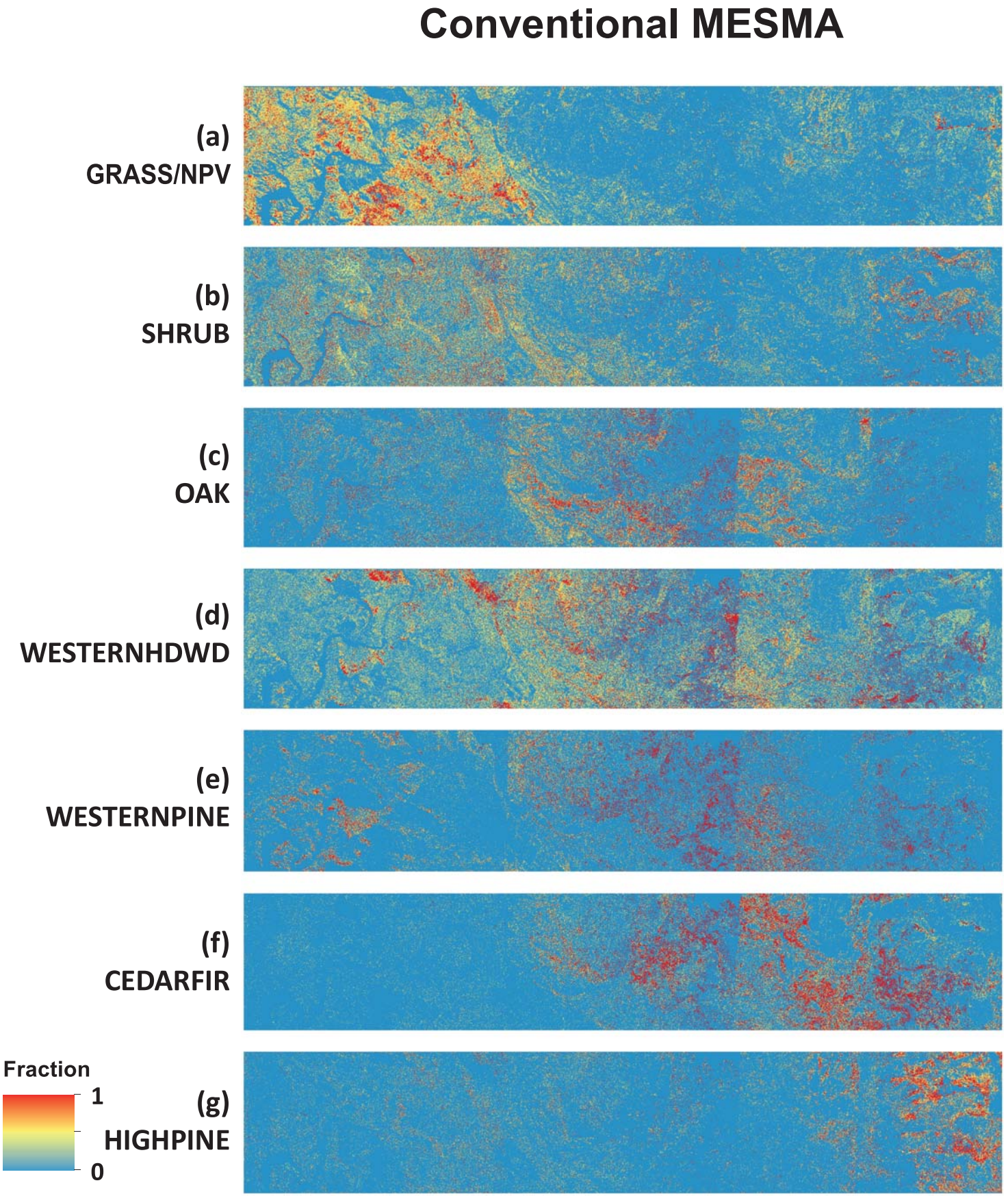


Figure 4

Click here to download Figure: Boganetal_Figure4.pdf

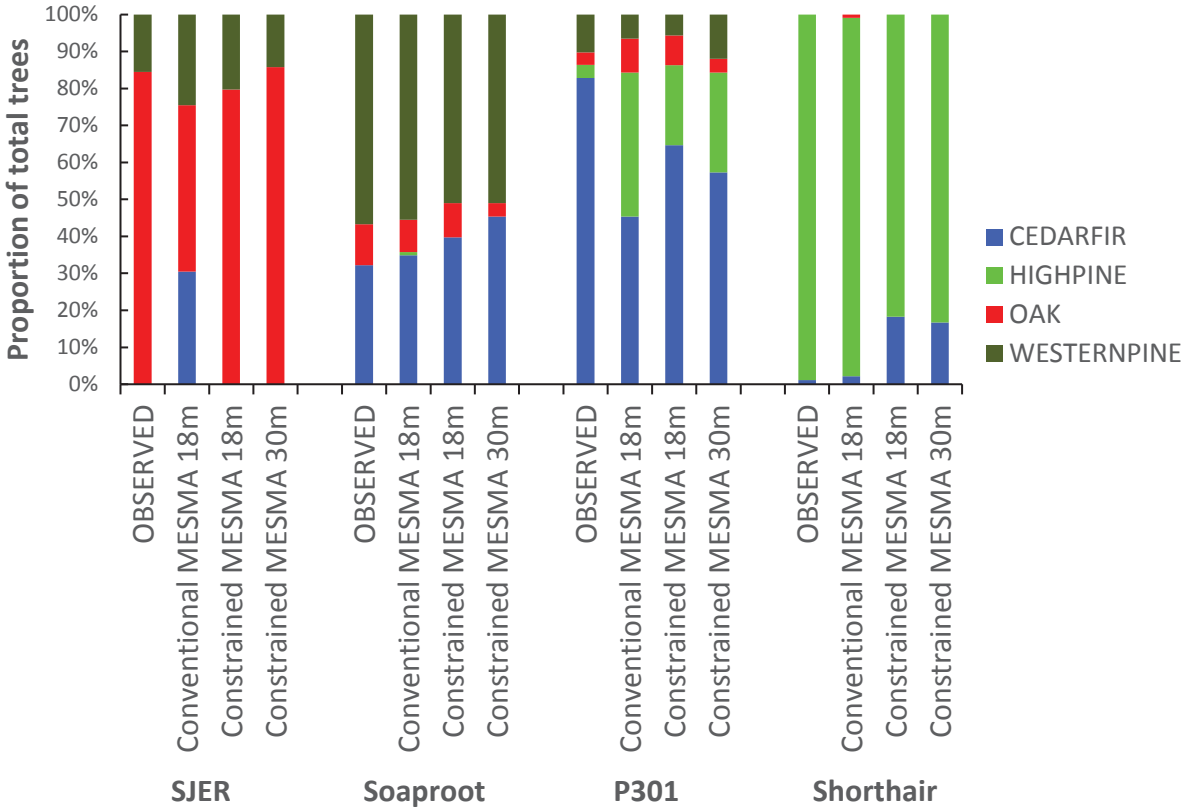


Figure 5

Click here to download Figure: Boganetal_Figure5.pdf

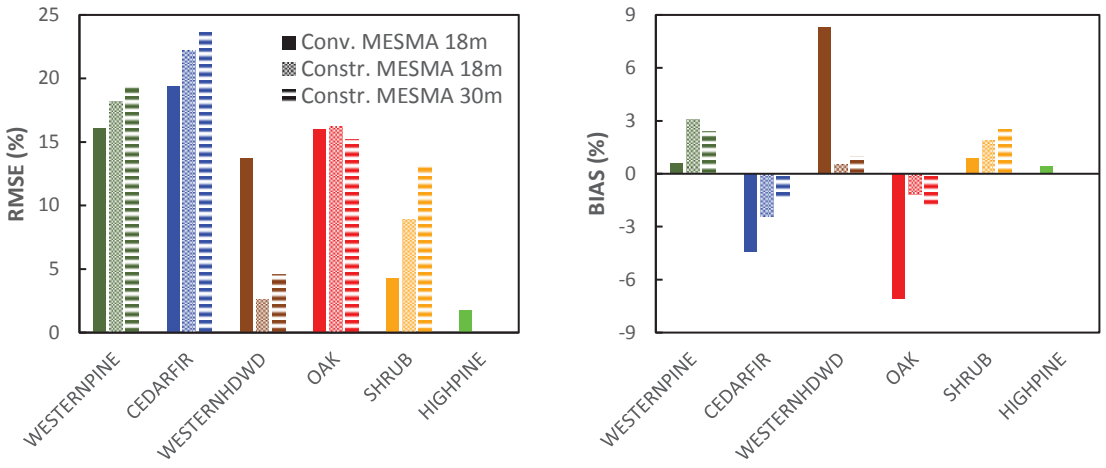


Figure 6
Click here to download Figure: Boganetal_Figure6.pdf

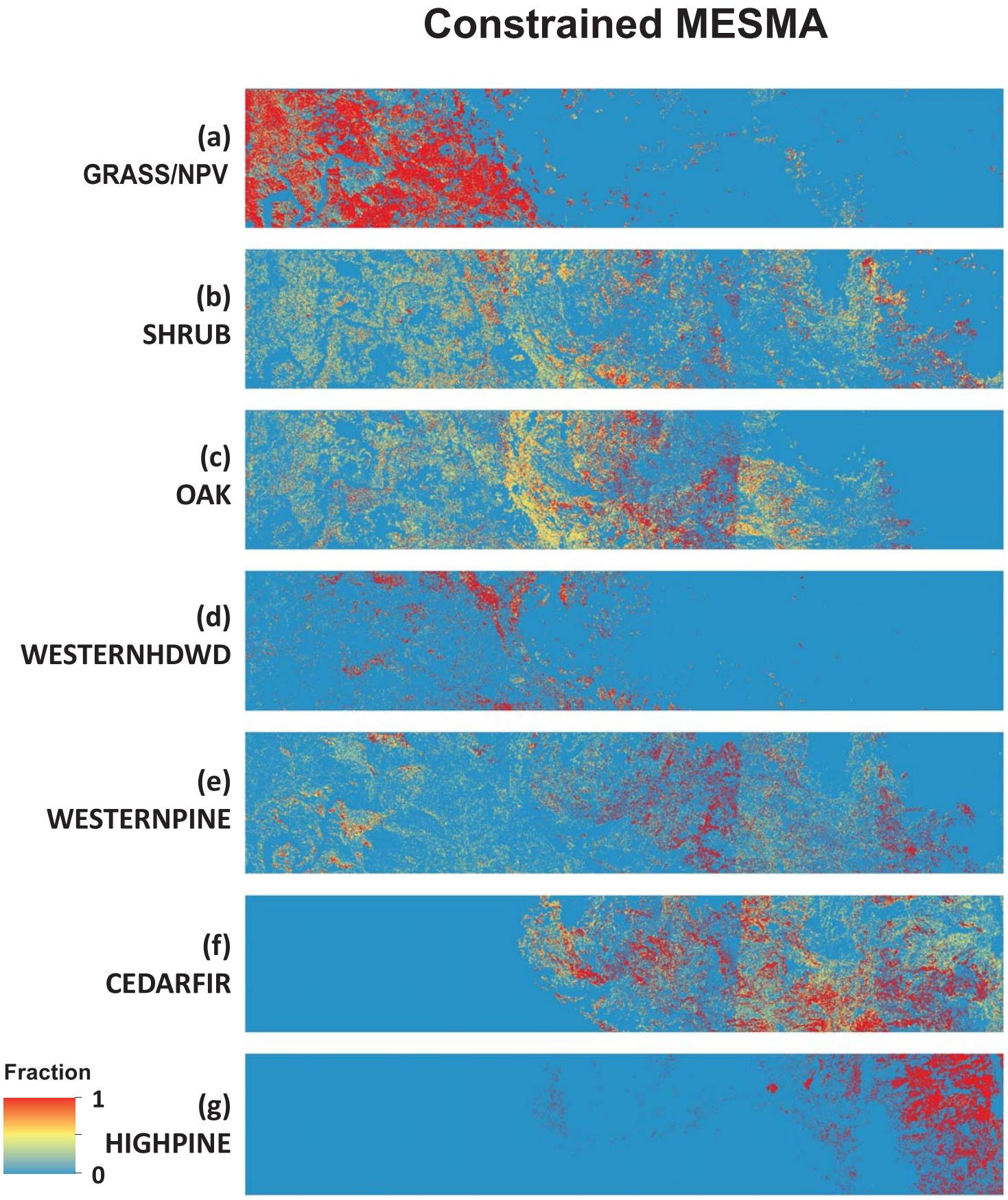


Figure 7

Click here to download Figure: Boganetal_Figure7.pdf

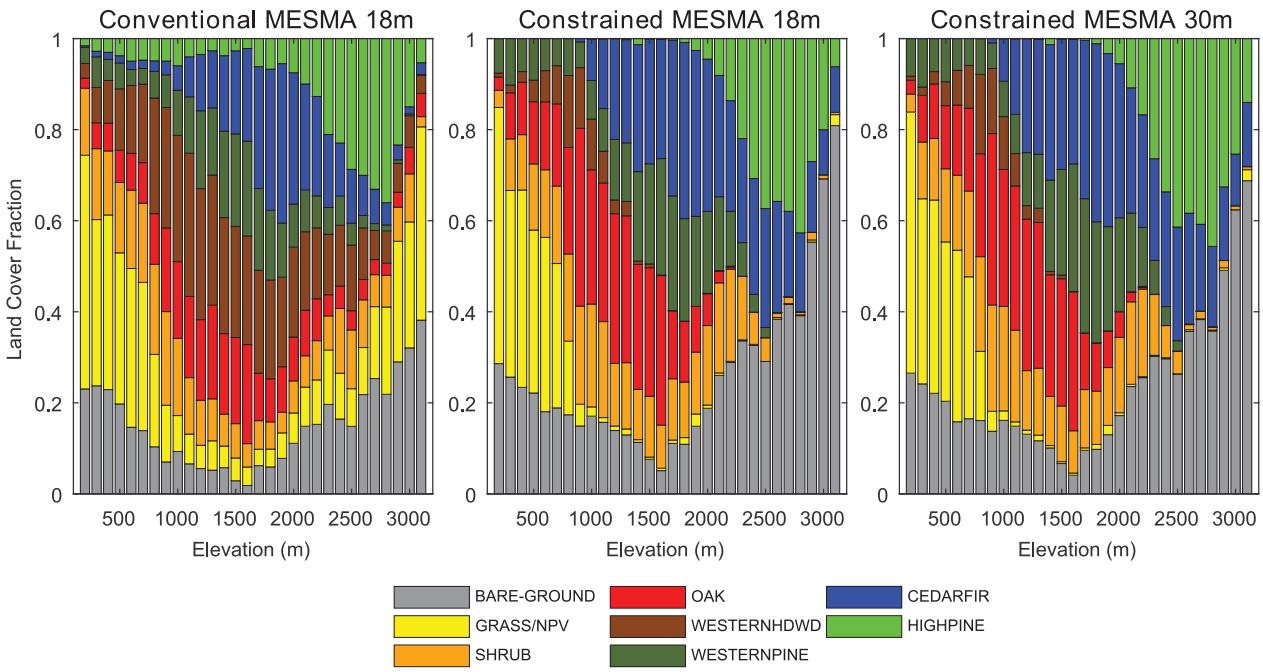


Figure 8

Click here to download Figure: Boganetal_Figure8.pdf

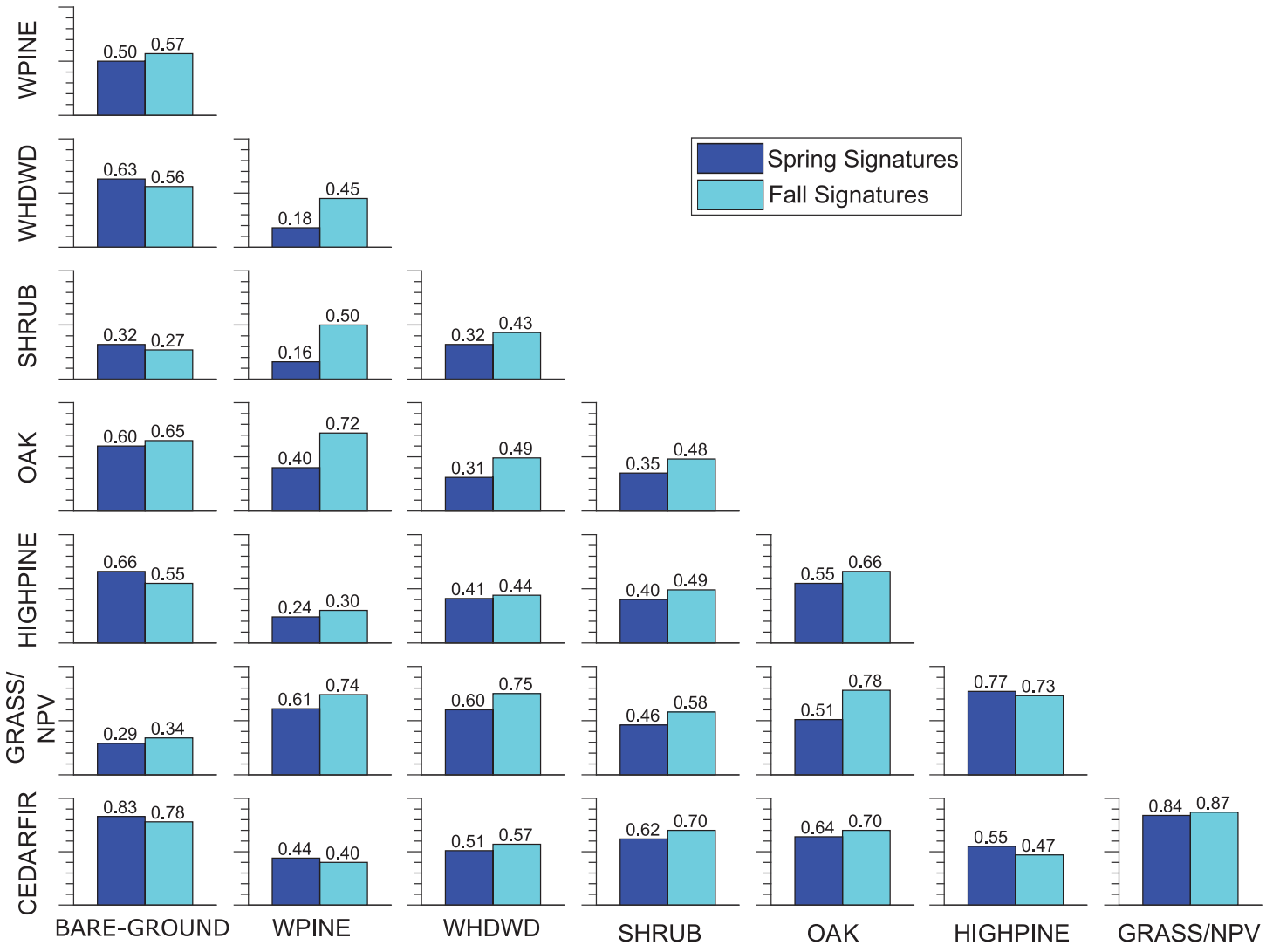


Figure 9

Click here to download Figure: Boganetal_Figure9.pdf